

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 5, 2004, 17:59:26 ; Search time 4223 Seconds  
(without alignments)  
5614.170 Million cell updates/sec

Title: US-09-943-857-4

Perfect score: 2864

Sequence: 1 SMNSRCPAGRLGSLVPTAKLA.....DNFRTAGYDALMTNPSFFV 547

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn2.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09943857/runat\_29072004\_090900\_29732/app\_query.fasta\_1.711  
-DB=GenEmbl -QEXT=fastcap -SUFFIX=rgc -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09943857@cgn\_1.1.2568 @runat\_29072004\_090900\_29732 -NCFU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.mu.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2671	93.3	1855	8	CCLIP3 X66006 C.cylindrac
2	2600	90.8	1647	8	AY464467 AY464467 Candida c
3	2513	87.7	1647	6	A81171 A81171 Sequence 1
4	2513	87.7	1688	12	AF044078 AF044078 Synthetic
5	2416	84.4	1857	8	CCLIPASE X16712 C.cylindra
6	2415	84.3	1733	6	A48377 A48377 Sequence 6
7	2415	84.3	1733	6	CCLIP1 X64703 C.cylindrac
8	2393	83.6	1605	6	A81172 A81172 Sequence 2
9	2392	83.5	1532	6	AX709921 AX709921 Sequence
10	2347	81.9	1812	8	CCLIP5 X66008 C.cylindrac
11	2268	79.2	1785	8	CCLIP4 X66007 C.cylindrac
12	2212	77.2	2043	8	CCLIP2 X64704 C.cylindrac
13	2034	71.0	1541	6	AX709925 AX709925 Sequence
14	2020	70.5	1548	6	AX709923 AX709923 Sequence
15	1903	66.4	1511	6	AX709927 AX709927 Sequence
16	1875	65.5	1469	6	AX709919 AX709919 Sequence
17	1107	38.7	1828	6	A74255 A74255 Sequence 11
18	1095	38.2	1635	8	GC002525 U02525 Geotrichum
19	1092	38.1	1692	6	E02678 E02678 cDNA encodi
20	1089	38.0	1635	6	A48376 A48376 Sequence 5
21	1089	38.0	1635	8	GC002387 U02387 Geotrichum
22	1088	38.0	1635	8	GC002622 AB000260 Geotrichu
23	1087	38.0	1783	8	AB000260 Geotrichu
24	1084	37.8	1635	8	GC002524 X81656 G.candidum
25	1081	37.7	1635	8	GCTAGL U02524 Geotrichum
26	1081	37.7	1635	8	GC002625 U02625 Geotrichum
27	1081	37.7	1692	8	GGELR X78032 G.geotrichu
28	1078	37.6	1674	6	E02497 E02497 cDNA encodi
29	1078	37.6	1767	8	GCALIP2 D0697 Geotrichum
30	1070	37.4	1635	8	GC002623 U02623 Geotrichum
31	1069	37.3	1635	8	GC002541 U02541 Geotrichum
32	938.5	32.8	2045	6	A45569 A45569 Sequence 7
33	938.5	32.8	2045	6	A47607 A47607 Sequence 7
34	938.5	32.8	2045	6	AR019333 AR019333 Sequence
35	801	28.0	4295	8	AY250996 AY250996 Aspergill
36	693	24.2	1934	8	AK107259 AK107259 Oryza sat
37	632	22.1	2000	8	ANGLAAS X56442 A.niger gla
38	536.5	18.7	13546	1	AE011751 AE011751 Xanthomon
39	495	17.3	1971	5	GG306928 XJ306928 Gallus ga
40	491.5	17.2	2490	5	TCACBR X03439 Torpedo cal
41	489	17.1	2357	5	TMACHE X05497 Torpedo mar
42	484	16.9	2033	3	AF080184 AF080184 Meloidogy
43	482	16.8	2133	3	AF075718 AF075718 Meloidogy
44	481	16.8	2089	6	AX306139 AX306139 Sequence
45	481	16.8	2089	10	MMACHE X56518 Mouse mRNA

# ALIGNMENTS

RESULT 1

111 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30







```

QY 331 AspMetTyrIysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350
Db 997 GACATGACGCTTGTGCGGAGGCAAGTATGCCAATCCCTGTGATCAICGGCGAC 1056
QY 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
Db 1057 CAGAACGACGAGGCACTTCTTTGGCACCAGCAGCTTGAACTGACCGATGCCAG 1116
QY 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
Db 1117 GCCCGCGAGTATTTCAAGCAGAGCTTTGTCCAGCCGACGACCGGAGATCGACGTTG 1176
QY 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
Db 1177 ATGACGCGCTACCCCGCGACATCACCCAGGCGAGCCGTTGCACACGGGTATTTCAAC 1236
QY 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
Db 1237 GCCTCACCCCGCAGTTCAAGAAATCAGCGCGGTCTCGGCGACCTTGGCTTTACGCTT 1296
QY 431 AlaArgTyrPheLeuAsnHisPheGlnGlyThrIysTyrSerPheLeuSerLys 450
Db 1297 GCTGTCGCTACTTCTCAACCACTACACCGCGCGCACCAAGTACAGTTCCTCAGCAAG 1356
QY 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
Db 1357 CAGCTCAGCGGCTTGCCTGCTCGGAACGTTCCACTCCAAACGACATTTGTTCCAGGAC 1416
QY 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAlaPheIleAlaPheAlaThrAsp 490
Db 1417 TACTTGTGGCAGCGGCTCGCTCATCAACACGCGTTCATCGCTTTCGTCACGAG 1476
QY 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db 1477 TTGGACCCCAACACCGCGGTTGTGTGAAGTGCCTGAGTACACAGCAGCACCCAG 1536
QY 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
Db 1537 AGCGGCAACAACCTGATGATGATCAACGCGCTTGGGCTTGTACACCGGCAAGCAACTTC 1596
QY 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db 1597 CGACCGCGGCTACGCGGTTGTCTCCAAACCGCGAGCTTCTTGTG 1647

RESULT 4
AF044078
LOCUS
DEFINITION
  Synthetic construct triacylglycerol hydrolase (lip1) gene, complete
  cds.
ACCESSION
  AF044078
VERSION
  1
KEYWORDS
  synthetic construct
  synthetic construct
  artificial sequences.
ORGANISM
  Brocca,S., Schmidt-Dannert,C., Lotti,M., Alberghina,L. and
  Schmid,R.D.
REFERENCE
  1 (bases 1 to 1688)
  Design, total synthesis and functional overexpression of the
  Candida rugosa lip1 gene coding for a major industrial lipase
  Protein Sci. (1998) in press
  2 (bases 1 to 1688)
  Brocca,S., Schmidt-Dannert,C., Lotti,M., Alberghina,L. and
  Schmid,R.D.
  Direct Submission
  Submitted (22-JAN-1998) Institute of Technical Biochem., University
  of Stuttgart, Altmannring 31, Stuttgart 70569, Germany
  Location/Qualifiers
  1..1688
  /organism="synthetic construct"
  /mol_type="genomic DNA"
  /db_xref="taxon:32630"
  /note="designed gene encoding lip1 of Candida rugosa"
FEATURES
  source

```

gene  
CDS

```

19..1668
/gene="lip1"
19..1668
/gene="lip1"
/EC_number="3.1.1.3"
/function="lipase"
/note="Lip1; codon optimized for expression in
Saccharomyces cerevisiae"
/codon_start=1
/product="triacylglycerol hydrolase"
/protein_id="AAC02093.1"
/db_xref="GI:2852390"
/translation="MELALSLIASVAAPATLANTITGLNAINAEAFGIPPA
EPVGNLRKDPVPSGSDGKFTSYSPSCMOQNEGYEENLPAALDLYMQSKVF
EAVPSSEDLATINVRPPTKAGANLPVLMIRGCGFEVGGTSTFPQAMITYKIAM
GXPLHVSVYVRVSWGFLAGDEIKAGSANAGLQDRLGMQWVAADIAFGGDPFKV
TIFGSAGSMVWCHILMNDGNTYKGFPRAGIMQSGAMVPSDAVDGIGYNEIFDL
IASNACGGSADKLACLRGSSDTLEATNPTFGFLAYSLSLSYLPDPDGVNITDLM
YALVRGKYANIPVIIGDQNDGTFGTSLNVTDAQAREYFKQSFVHASDAEIDTL
MTAYPGDITQGGPFDITNALTPQKRISAVLGDGLFTLARRYFLNHYTGGTKYSEL
SKOLSLPLVLTGTFHSNDIVPDYLLGSSGLIYNNAFIAFATDLDPTAGLLVKNPEYT
SSSQSNLMLMINALGLYTGKDNFRTAGYDALFSPNPPFFV"

```

## ORIGIN

Alignment Scores:

Pred. No.:	1..66e-201	Length:	1688
Score:	2513.00	Matches:	470
Percent Similarity:	93.48%	Conservative:	32
Best Local Similarity:	87.52%	Mismatches:	35
Query Match:	87.74%	Indels:	0
DB:	12	Gaps:	0

US-09-943-857-4 (1-547) x AF044078 (1-1688)

QY 11 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30  
 Db 55 GTTGTCTGTCGCCCAACCGCCACTTTTGGCTAACCGTGACACCATCACCGGTTTGAACGCC 114  
 QY 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArg 50  
 Db 115 ATCATCAAGAGCCCTTCTTGGGTATTCCATTGGCGAACCCACCATCGTGGTAACATTGAGA 174  
 QY 51 PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly 70  
 Db 175 TTCAAGGACCCAGTTCCTACTCCGTTCTTGGATGGTCAAAAGTTCACTCTTACCGT 234  
 QY 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90  
 Db 235 CCATCTGTATGCAACAAACCCAGAGGTACCTAGCAAGAAACTTGCACAAAGGCGAGCT 294  
 QY 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110  
 Db 295 TTAGATCTGGTTATGCAATCCAAAGTTTTCGAAGCTGTTTCTCCATCTTCTGAAGACTGT 354  
 QY 111 LeuThrIleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeuProValMet 130  
 Db 355 TTGACCATTAAGTTGTAGACCACCCGCGACAAAGGCTGGTCCCACTTGCAGGTATG 414  
 QY 131 LeuTrpIlePheGlyGlyGlyPheGluIleGlySerProThrIlePheProAlaGln 150  
 Db 415 TTGTGGATCTTTGGTGGTGGTTTGAAGTTGGTGTACTAGTACCTTCCTCCAGCCCAA 474  
 QY 151 MetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArg 170  
 Db 475 ATGATTACCAAGCTTATTGCTATGGGTAGGCAATCATCCAGCTTCTTGTCACATACAGA 534  
 QY 171 ValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAla 190  
 Db 535 GTCTCAGCTGGGTTTCTTGGTGTGACGAAATCAAGGCCAAGGTTCTGCCAACGCC 594  
 QY 191 GlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly 210  
 Db 595 GGTTTGAAGGACCAAGATTGGGTATGCAATGGGTGGCTGACCAACATTGCTCTTTGCT 654

QY 211 GlyAspProSerLysValThrIlePheGlySerAlaGlySerMetSerValLeuCys 230  
 Db 655 GGTGATCAACTAAGGTACTATCTTTGGTGAATCTCTGCTTCTAATGTCGTCATGTGT 714  
 QY 231 HisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250  
 Db 715 CACATTTTGGACGAGCGGTGACACACTTACAGGGTAAGCCATTTTCAGAGCTGGT 774  
 QY 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270  
 Db 775 ATCATGCAATCTGGTGTATGTTTCCATCTGACCGCTGACGGTATCTACGGTAACGAA 834  
 QY 271 IleTyrAspLeuPheValSerAlaGlyCysGlySerAlaAspGlySerLeuAlaCys 290  
 Db 835 ATTTTGTACCTTTGGTGTCCAAAGCTGTGTGCTCTGCTCTGCAAGTTGGCTGT 894  
 QY 291 LeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310  
 Db 895 TTGAGAGGTGTTCCTCTGACACTTTTGAAGACGCCACCAACACACCCCTGGTTCTTG 954  
 QY 311 AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330  
 Db 955 GCTTACTCTCTTAAGATTCTTCTTGTGATCTTCTTGTGCAAGACGAGCGGTAAACATCACCGAC 1014  
 QY 331 AspMetTyrLysLeuValAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350  
 Db 1015 GACATGTACGCTTTGGTTAGAGAGGTAAAGTATGCCACATCTCTGTATCATCGGTGAC 1074  
 QY 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrAsnAlaGln 370  
 Db 1075 CAAACGACGAAGTACCTTCTTGTGATCTTCTTCTTGTGAAGCTTACCACCTGATGCCAA 1134  
 QY 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390  
 Db 1135 GCCAGAGAATATTTCAAGCAATCTTTGTCCACGCTACGCGTGAATCGACACTTTG 1194  
 QY 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410  
 Db 1195 ATGACTGTCTACCCAGGTGATCATCTCAAGTTCTCCATTGACACTGGATTCCTAAC 1254  
 QY 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430  
 Db 1255 GCCTTGACCCACCAATTCAGAGAACTCTCTGCTGTTTGGGTGACTTGGGTTTACTTTG 1314  
 QY 431 AlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLys 450  
 Db 1315 GCTCGTAGATATCTTGAACCACTACACCGGTGATCCCAAGTACTCTTCTTGTCTAAG 1374  
 QY 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470  
 Db 1375 CAATTGTCTGGTTTCCAGTTTGGGTATCTTCCACTCCACGATATGCTTCCACGAC 1434  
 QY 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490  
 Db 1435 TACTTGTGGTCTCTGTTCTTGTATCTACAAACGCTTTCATTGCTTTTGCCACTGAC 1494  
 QY 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerGln 510  
 Db 1495 TTGGACCCAAACACCGCGGTGTGTGGTTAAGTGGCCAGAAATACACCTCTTCTTCTCAA 1554  
 QY 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530  
 Db 1555 TCTGTATACACTTGATGATGATCAAGCTTTGGGTGTGATACCCCGTAAGGACAACTTC 1614  
 QY 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547  
 Db 1615 AGAACCGCGGTACGACGCTTGTCTCCAAACCCACCATCTTCTTCTTGT 1665

## RESULT 5

CCLIPASE

LOCUS

DEFINITION

CCLIPASE 1657 bp mRNA linear PLN 12-DEC-1997  
 C. cylindracea mRNA for lipase I, partial cds.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

X16712  
 X16712.1 GI:2548  
 lipase; lipase I.  
 Candida cylindracea  
 Candida cylindracea

## REFERENCE

1 (bases 1 to 1657)

## AUTHORS

Kawaguchi, Y., Honda, H., Taniguchi-Morimura, J. and Iwasaki, S.

## TITLE

The codon CUG is read as serine in an asporogenic yeast *Candida*

## JOURNAL

*cylindracea*

## MEDLINE

Nature 341 (6238), 164-166 (1989)

## PUBMED

89384874

## FEATURES

2506450

## source

Location/Qualifiers

1..1657

/organism="Candida cylindracea"

/mol\_type="mRNA"

/strain="MS-5"

/db\_xref="taxon:44322"

/clone="lambda CL115"

<1..1657

/note="for lipase 1"

<1..1617

/note="Protein sequence is in conflict with the conceptual

translation"

/codon\_start=1

/codons=(seq:"ctg",aa:Ser)

/product="lipase 1"

/protein\_id="CAA34684.1"

/db\_xref="GI:2683215"

/db\_xref="GOA:P20261"

/db\_xref="SWISS-PROT:P20261"

/translation="SVAAAPTATLANGDTITGLNAINAEFLGIPAPPPVGNLRFKD

PVPVSGSLDGKQKTSYGPSCQONPECTYBENLKAALDLVMQSKVEAVSPSSDCL

TINVRPPTGKAGANLPMWLKFGFEGVGTSTFPAQMLTKSIAMGKPIIHVSVNY

RVSNGFLAGDEIKAESANGLKQRLGMQWADNIAAEGDPTKVTIFGESAGMS

VMCHLNWDGNTYKPLFRAGIMQSGAMVPSDAVDGIYGNIEFLDLASNAGGSSAS

DKLACLRGVSDDTLEDATNTTPGLAYSLSRLSYLPRPDGNIITDDMYALVREGKIAN

IPVTIGQNDGTFPGTSSLNVTDAQAREYFKQSFVHASDAEITDMLTAYPDITQG

SPDPTGLNALTPOFKRISAVIGDLGFTLARRYFLNHYTGGTKTSFLSKQLSGPLVLG

TFHSDIVFDQYLLGSGSLIYNNAFIAFATDLDPTNAGLLVKWPEYTTSSSQSGNNLNM

INAGLYTGKDNFTAGYDALFSNPPSFV"

<1..12

13..1614

/product="lipase 1"

## ORIGIN

sig\_peptide

mat\_peptide

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-943-857-4 (1-547) x CCLIPASE (1-1657)

QY 11 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30

Db 4 GTGGTGTGTCGCCCCACCGCCACGCTCCGCCAACGGACACCATCACCGGTCTCAACGCC 63

QY 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArg 50

Db 64 ATCATCAACGAGCGGCTTCTCGGCATTCCTTTCGAGCGCGCGCTCGGCAACCTCCGC 123

QY 51 PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly 70

Db 124 TTCAGAGCCCCCGTCCGCTACTCCGCTCGCTCGATGCGGACCAAGTTACGCTGTACGCG 183

QY 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90

Db 184 CGCTGTGATGACAGCAGACAGACCCCGAGGACCTACGAGGAGAACCTCCCAAGGACGG 243

QY 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110  
 Db 244 CTGACTTGGTGTGATGTCAGTCCAAAGTGTCTTGGAGCGGTGCTGCCGCTGAGGAGACTGT 303  
 QY 111 LeuThrIleAsnValValArgProProGlyThrIysAlaGlyAlaAsnLeuProValMet 130  
 Db 304 CTACCAATCAACGTGGTGGCGCGCCGCGGACCAAGGCGGGTGCACCTCCCGGTGATG 363  
 QY 131 LeuTriPheGlyGlyPheGluIleGlySerProThrIlePheProAlaGln 150  
 Db 364 CTCTGGATCTTGGCGCGGGTGTGGGTGGGTGGCCACGACCTTCCCTCCCGCCAG 423  
 QY 151 MetValThrIysSerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArg 170  
 Db 424 ATGATCACCAGAGCATGTCATGGGCAAGCCCATCATCAGCGTGAGCGTCACTACCGC 483  
 QY 171 ValAlaSerTrpGlyPheLeuAlaGlyAspIleIysAlaGluGlySerGlyAsnAla 190  
 Db 484 GTCTCGTGGTGGGGTCTTGGCTGGCGAGCATCAAGGCCGAGGCGCATGTCACAGCC 543  
 QY 191 GlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAlaAspAsnIleAlaGlyPheGly 210  
 Db 544 GGTTCAGAGCAGCGCTTGGCATGTCAGTGGTGGCGGACCACTTGGCGGTTCGGC 603  
 QY 211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230  
 Db 604 GGCAGCCCGACCAAGGTGACCATCTTGGCGAGCTGGCGGCGAGCATGTCGGTTCATGTGC 663  
 QY 231 HisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250  
 Db 664 CAAITCTCTGGAAACGACGCGCAACACGTACAAGGGCAAGCGCTCTTCGCGGGGC 723  
 QY 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270  
 Db 724 ATCATGAGCTGGGGCCATGGTGGCGTGGACCGCTGGACGGCATCTACGGCAACGAG 783  
 QY 271 IleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290  
 Db 784 ATCTTTGACTCTTGGCGTTCGAACGCGGGTGGCGGCGAGCGCCAGCAGCTTGGGTGC 843  
 QY 291 LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310  
 Db 844 TTGCGCGGTGTGTGAGCGGACACGTGTGGAGGACGCCCAACCAACACCCCTGGGTCTTG 903  
 QY 311 AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330  
 Db 904 GCGTACTCTCTGTTGGTGTGTGTAACCTCCCGCGCGCCGACGCGGTGAACATCACCGAC 963  
 QY 331 AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350  
 Db 964 GACATGTAGCCCTTGGTGGCGGAGGCAAGTATGCCAACATCCCTGTGATCATCGCGGAC 1023  
 QY 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370  
 Db 1024 CAGAACGACGAGGGACCTCTTTGGACCCCTGCTGTGAACGTGACCAAGTACCGCCAG 1083  
 QY 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390  
 Db 1084 GCCCGGAGTACTTCAAGCAGCTGTGTTCACCGCCAGCGACGCGGAGATCGACAGTTG 1143  
 QY 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410  
 Db 1144 ATGACGGCGTACCCCNNGACATCACCCAGGGCGCTCGGTTCGACACGGGTATCTCAAC 1203  
 QY 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430  
 Db 1204 GGCCTCACCCCGAGTTCAGAGAAATCTCTGGGGGTGCTCGCGGACCTTGGCTTTAGCGTT 1263  
 QY 431 AlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLys 450  
 Db 1264 GCTCTGCTGCTACTTCTTCAACCACTACACCGCGGCGCACCAAGTACTCTCTCTCTGAAG 1323

QY 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470  
 Db 1324 CAGCTCTCGGNTTGGCGGTGCTCGGAACTGTCATCTCCAAACGACATGCTTCCAGGAC 1383  
 QY 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490  
 Db 1384 TACTTGTGGCGACGCGCTCGCTCATCTACAACAACGCGTTCATTGGCTTGGCACGGAC 1443  
 QY 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510  
 Db 1444 TTGAGCCCCAACCCGCGGGTGTGTGGTGAAGTGGCCGAGTACACACGAGCGCTCGAG 1503  
 QY 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530  
 Db 1504 CTGGGCAACAACTTGATGATGATCAACGCTTGGCTTGTACACCGGCAAGGACAACTTC 1563  
 QY 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547  
 Db 1564 CGCACCGCGGCTACGACGCTGTCTTCTCCAAACCGCGCGCTGCTTCTTTGTG 1614  
 RESULT 6  
 A48377  
 LOCUS A48377 1733 bp DNA linear PAT 07-MAR-1997  
 DEFINITION Sequence 6 from Patent WO9603511.  
 ACCESSION A48377  
 VERSION A48377.1 GI:2302168  
 KEYWORDS Candida rugosa  
 SOURCE Candida rugosa  
 ORGANISM  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 REFERENCE 1 (bases 1 to 1733)  
 AUTHORS Alibert, G., Moulounqui, Z. and Boudet, A.  
 TITLE METHOD FOR PRODUCING FATTY ACIDS OR DERIVATIVES THEREOF FROM OIL  
 JOURNAL PLANT  
 PATENT: WO 9603511-A 6 08-FEB-1996;  
 TOULOUSE INST NAT POLYTECH (FR)  
 COMMENT Other publication AU 2984995 960222  
 Other publication FR 2722798 960126.  
 FEATURES  
 source  
 1. .1733  
 /organism="Candida rugosa"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:5481"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,94e-193 Length: 1733  
 Score: 2415.00 Matches: 454  
 Percent Similarity: 90.50% Conservative: 32  
 Best Local Similarity: 84.54% Mismatches: 51  
 Query Match: 84.32% Indels: 0  
 DB: Gaps: 0  
 US-09-943-857-4 (1-547) x A48377 (1-1733)  
 QY 11 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30  
 Db 37 GTGGTGTCTGCCCGCCACCGCCACGCTCGCCAAAGCGGACCATCACCGGTCTCAACGCG 96  
 QY 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArg 50  
 Db 97 ATCATCAACGAGCGCTTCTCTGGCATTCCTTTGCGAGCGCGCGGTGGGCAACCTCCG 156  
 QY 51 PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly 70  
 Db 157 TTCAAGGACCCCGTCCCGTACTCCGCTCGCTCGATGCGCAGAGTTCACGCTGTACGCG 216  
 QY 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90  
 Db 217 CGCTGTGTCATGCACGAAACCCCGAGGCGACCTACGAGGAGAACCTCCCCAAGCAGCG 276  
 QY 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110

```

Db      277  CTGAGCTTGGTGTGAGTCCAGGTCCTTGAAGCGGTGTGCGCTGAGGAGGACTGT 336
Qy      111  LeuThrIleAsnValValArgProGlyThrIlysAlaGlyAlaAsnLeuProValMet 130
Db      337  CTCACCATCAAGTGGTGGCGCGCGGCAACCAAGCGGGGTGCCAACCTCCCGGTGATG 396
Qy      131  LeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePheProAlaGln 150
Db      397  CTCGTGATCTTGGCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 456
Qy      151  MetValThrIlySerValLeuMetGlyIleHisIleIleHisValAlaValAsnTyrArg 170
Db      457  ATGATCACCAGACGATGGCATGGGCAAGCCATCATCCACGAGGAGGTCAACTACCGC 516
Qy      171  ValAlaSerTrpGlyPheLeuAlaGlyAspAspIleIysAlaGluIysSerGlyAsnAla 190
Db      517  GTGTCGTCTGGTGGGCGTCTTGGTGGGACGAGATCAAGGCGGAGGCGAGTGCACAGCC 576
Qy      191  GlyLeuIysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly 210
Db      577  GGTGTGAAGGACCGAGCTTGGGATGCGATGCGTGGTGGTGGTGGTGGTGGTGGTGGT 636
Qy      211  GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValIleuCys 230
Db      637  GGCACCGCACCAAGTGACCATCTTGGCGAGCTGGCGGCGGCGGCGGCGGCGGCGGCG 696
Qy      231  HisLeuIleTrpAsnAspGlyAspAsnThrTrpIysGlyLysProLeuPheArgAlaGly 250
Db      697  CACATCTCTGGAACGAGCGGACCAACACGTACAGGCAAGCGGCTCTTCGCGCGGCGG 756
Qy      251  IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270
Db      757  ATCATGACGCTGGGCGCATGTCGCTGGACGCGGTGGACGGCATCTACGGCAACGAG 816
Qy      271  IleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290
Db      817  ATCTTTGACCTTGGCGTGGACCGCGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCG 876
Qy      291  LeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
Db      877  TTGCGCGGTGTCTGAGCGACAGTGTGGAGGACGCCACCAACACCGCCCTGGGTCTTG 936
Qy      311  AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330
Db      937  GCGTACTCTCTGCGGTGTCTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 996
Qy      331  AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350
Db      997  GACATGTACGCTTGGTGGCGGAGGCAAGTATGCCAACATCCCTGTGATCATCGCGGAC 1056
Qy      351  GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
Db      1057  CAGAACGACGAGGCGACCTTCTTGGCACCCCTGCTGTGAACGTGACACCGGATGCCAG 1116
Qy      371  AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
Db      1117  GCCCGGAGTACTTCAAGCAGCTGTGTGTCACGCGGCGGCGGCGGAGATCGACAGTTG 1176
Qy      391  MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
Db      1177  ATGACGGCGTACCCCGGCGACATCAACCGGCGCTGCTCGGCGACCTTGGCTTTACGCTT 1236
Qy      411  AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
Db      1237  GCGCTCACCCCGAGTTCGAAGAGATCCCTGGCGGTGCTCGGCGACCTTGGCTTTACGCTT 1296
Qy      431  AlaArgTyrPheLeuAsnHisPheGlnGlyThrIysTyrSerPheLeuSerLys 450
Db      1297  GTCGTGCGTACTTCTCTCAACCACTACACCGGCGGCGGCGGAGTACTCATTCCTCTGAAG 1356
Qy      451  GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470

```

```

Db      1357  CAGCTCCTGGCTGGCGGTGCTCGGAACGTTCCACTCCACGACGATCTGCTTCAGGAC 1416
Qy      471  TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
Db      1417  TACTTGTGGGAGCGGCTGCTCATCTACACAAACGCGTTTCAATGGTTTGCCACGAC 1476
Qy      491  LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db      1477  TTGACCCCAACACCGCGGGTGTGGTGAAGTGGCCGAGTACACACGAGCGCTCGAG 1536
Qy      511  SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
Db      1537  CTGGGCAACAACTTGATGATGATCAACGCTTGGGCTTGATACCCGCAAGGACACTTC 1596
Qy      531  ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db      1597  CGCACCGCGGTACGAGCGGTGTCTCTCAACCCGCGCTGTTCTTTGTG 1647

RESULT 7
CCLIP1      1733 bp      DNA      linear      PLN 27-APR-1993
LOCUS      C.cylindracea LIP1 gene for lipase.
ACCESSION  X64703
VERSION     X64703.1 GI:2544
KEYWORDS    Lip1 gene.
SOURCE      Candida cylindracea
ORGANISM    Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1 (bases 1 to 1733)
AUTHORS     Longhi,S., Fusetti,F., Grandori,R., Lotti,M., Vanoni,M. and
            Alberghina,L.
TITLE       Cloning and nucleotide sequences of two lipase genes from Candida
            cylindracea
JOURNAL     Biochim. Biophys. Acta 1131 (2), 227-232 (1992)
MEDLINE     92305068
PUBMED      1610906
REFERENCE   2 (bases 1 to 1733)
AUTHORS     Longhi,S.
TITLE       Direct Submission
JOURNAL     Submitted (28-FEB-1992) S. Longhi, Universita' degli Studi di
            Milano, Via Celoria 26, 20133 Milano, ITALY
COMMENT     See also x64704
            X66006, X66007 and X66008 are related sequences in the description
            of C.cylindracea Lip1 and Lip2 sequences (x64703, x64704).

FEATURES
            Location/Qualifiers
                1..1733
                    /organism="Candida cylindracea"
                    /mol_type="genomic DNA"
                    /strain="ATCC 14830"
                    /db_xref="taxon:44322"
                    /clone="A"
                1..1650
                    /gene="LIP1"
                1..1650
                    /gene="LIP1"
                    /EC_number="3.1.1.3"
                    /codon_start=1
                    /product="triacylglycerol lipase"
                    /protein_id="CAA45957.1"
                    /db_xref="GI:1325987"
                    /db_xref="GOA:P20261"
                    /db_xref="SWISS-PROT:P20261"
                    /translation="MELALLLITASVAAAPTATLANGDTITGLNAINIENAPFLGPFA
                    EPPVGNLRFKDPVPYVSGSDGQKFTLYGLCMQNPGETYIENLPAALDLVMQSKVF
                    EAVLPLSDCLTINVRPPTGKAGANLPMWIMFGGGEVGTSTFFPAQMTKSIAM
                    GKPLIHVSNVFVSWGFLAGDEIKAEGSANAGLKQRLGMOWADVADVIGYNEIFDL
                    TIFGLAAGSMVCMCHLNDGNTYKGLPEAGIMOLGAMVPLDVGIVGNEITDM
                    IASNAGCGSADKLACLKGLSDTLEADNTNNTPGFLAYSSLRLLYLPRPDGVNITDM
                    VALVRKYANTIPVIGQNDDEGTFFGLLNLTVDQAQREYFKQLFVHASDAEIDTL
                    MTPVPGDITQGLTGHFNNDIVFDYLLGSGSLIYNNAFIAPATDLDENTAGLLVKVPEYT
                    LKQLGLGLVGTGFTSNDIVFDYLLGSGSLIYNNAFIAPATDLDENTAGLLVKVPEYT
                    SSLQLGNLMLNALGLYTGKDNFRTAGDYDALFSNPPLFFV"

```

```

sig_peptide      1. .45
                  /gene="LIP1"
mat_peptide      /note="triacylglycerol lipase"
                  46. .1647
                  /gene="LIP1"
                  /product="triacylglycerol lipase"
                  /EC_number="3.1.1.3"

ORIGIN
Alignment Scores:
Pred. No.:      2,945-193      Length:      1733
Score:          2415.00      Matches:     454
Percent Similarity: 90.50%      Conservative: 32
Best Local Similarity: 84.54%      Mismatches:  51
Query Match:    84.32%      Indels:      0
DB:             8      Gaps:         0

US-09-943-857-4 (1-547) x CCLIP1 (1-1733)

QY 11 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30
   :::::
Db 37 GTGGTGTCTGCTCCGCCACCGCTCGCCACCGGACACCATCACCGGTCTCAACGCC 96

QY 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArg 50
   :::::
Db 97 ATCATCAACGAGCGGTCTCTCGGCATTCTCGGCATTCTGCGAGCGCGGTGGGCAACCTCCGC 156

QY 51 PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly 70
   :::::
Db 157 TTCAGGACCCCGTCCCGTACTCCGGCTCGCTGATGGCCAGAAGTTACGCTGTACGCG 216

QY 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90
   :::::
Db 217 CCGCTGTGATGCAGCAGACACCGAGGCGCACTACAGAGAGAACCTCCCAAGGCGAGCG 276

QY 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110
   :::::
Db 277 CTCGACTTGTGATGCAGTCCAAAGGTGTGAGGCGGTGCTCCGCTGAGCGAGGACTGT 336

QY 111 LeuThrIleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeuProValMet 130
   :::::
Db 337 CTCACATCAACGTGTGCGCGCGCGCGGACCAAGCGGGTGCCAACTCCCGGATG 396

QY 131 LeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePheProProAlaGln 150
   :::::
Db 397 CTCGTGATCTTTGGCGCGGGTTGAGTGGGTGGCCACCGACCTTCCCTCCCGCCAG 456

QY 151 MetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArg 170
   :::::
Db 457 ATGATCACCAAGAGCATTTGCCATGGGCAAGCCCATCATCCAGTGTAGCGTCAACTACCGC 516

QY 171 ValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAla 190
   :::::
Db 517 GTGTGTGTGGGGGTCTTGCTGGCGACAGATCAAGCGCGAGGCGAGTCCCAACGCC 576

QY 191 GlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAsnIleAlaGlyPheGly 210
   :::::
Db 577 GGTTTGAAGGACAGCGCTTGGCATGCGAGTGGGTGGCGGACAACTTGGCGCGTTTGGC 636

QY 211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230
   :::::
Db 637 GCGGACCCGACCAAGGTGACCATCTTTGGCGAGCTGGCGGCGAGCATGTGCGTCTATGTC 696

QY 231 HisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250
   :::::
Db 697 CACATTCTGTGAACGACGCGCAACACGTAACAGGGCAACCGCTCTTCGCGCGGCG 756

QY 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270
   :::::
Db 757 ATCATGCAGCTGGGGCCATGTGTCGCTGGACGCGGTGACGGCATCTACGGCAACGAG 816

QY 271 IleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290
   :::::

```

```

Db 817 ATCTTTGACCTCTTGGCGTCGAACCGCGGCTGCGGAGCGCCAGCAGCAAGCTTGGCTGC 876
QY 291 LeuAqSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
   :::::
Db 877 TTGCGCGGTGTGCTGAGCGACAGCTTGGAGAGCGCCCAACCAACACCCCTTGGGTCTTG 936
QY 311 AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330
   :::::
Db 937 GCGTACTCTCGTTGCGGTGTGTGTACTCTCCCGCGCGCGCGCTGAACATCACCGAC 996
QY 331 AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleLeuGlyAsp 350
   :::::
Db 997 GACATGTACGCTTGTGTGCGGAGGCAAGTATGCCAACATCCCTGTGATCATCGCGGAC 1056
QY 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
   :::::
Db 1057 CAGACGACGAGGCGACCTTCTTTGGCACCTCTGTGTGAACGTGACCGAGATGCCAG 1116
QY 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
   :::::
Db 1117 GCCCGCGAGTACTTCAAGCAGCTGTGTGTCACGCGCGAGCGCGAGATCGACAGTTG 1176
QY 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
   :::::
Db 1177 ATGACGCGGTATCCCGCGGACATCACCCAGGCGCTGCGGTTCGACACGGGTATCTCAAC 1236
QY 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
   :::::
Db 1237 GCCCTCACCCCGAGTTCAGAGAAATCTTGGCGGTGCTCGCGGACCTTGGCTTTAGCTT 1296
QY 431 AlaArgTyrTyrPheLeuAsnHisPheGlnGlyGlyThrLysTyrSerPheLeuSerLys 450
   :::::
Db 1297 GCTGTGTGCTACTTCTCAACCACTACACCGCGCGGCAAGTACTCATCTCTCTGAG 1356
QY 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
   :::::
Db 1357 CAGCTCTGGGCTTGGCGGTGCTCGGAAGTTCCTCAACAGCATTTCTTCCAGGAC 1416
QY 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
   :::::
Db 1417 TACTTTGGTGGCGAGCGGTCTGCTCATCTACAAACAGCGGTTCATTGCGTTTGCACGGAC 1476
QY 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerGln 510
   :::::
Db 1477 TTGAGCCCAACACCGCGGGTGTGTGGTGAAGTGGCGGAGTACACCGAGCGCTGCAG 1536
QY 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
   :::::
Db 1537 CTGGGCAACAACTTGATGATGATCAACGCTTGGGCTGTGTACCGGCGAAGGACAACCTTC 1596
QY 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
   :::::
Db 1597 CGCACCGCGGTACAGCGGTGTCTTCCAAACCCCGCGCTGTCTTCTTGTG 1647

RESULT 8
A81172
LOCUS      A81172              1605 bp      DNA      linear      PAT 21-JAN-2000
DEFINITION Sequence 2 from Patent WO9914338.
ACCESSION A81172
VERSION    A81172.1  GI:6731631
KEYWORDS   .
SOURCE     unidentified
           unidentified
           unclassified.
REFERENCE  1 (bases 1 to 1605)
AUTHORS   Schmidt-Dannert,C. and Schmid,R.
TITLE     TOTAL SYNTHESIS AND FUNCTIONAL OVEREXPRESSION OF A CANDIDA RUGOSA
          LIP1 GENE CODING FOR A MAJOR INDUSTRIAL LIPASE
JOURNAL   Patent: WO 9914338-A 2 25-MAR-1999;
          SCHMIDT DANNERT CLAUDIA (DE); SCHMID ROLF (DE)
FEATURES   Location/Qualifiers
            1..1605
            /organism="unidentified"

```

/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"

## ORIGIN

## Alignment Scores:

Pred. No.: 1,87e-191 Length: 1605  
Score: 2393.00 Matches: 458  
Percent Similarity: 90.71% Conservative: 30  
Best Local Similarity: 85.13% Mismatches: 35  
Query Match: 83.55% Indels: 16  
DB: 6 Gaps: 2

US-09-943-857-4 (1-547) x A81172 (1-1605)

```

Qy 11 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30
Db 37 GTTGCTGCTGCCCAACCGCCACCTTTGGCTAAACGGTGACACCATCAACCGGTTGAACGCC 96
Qy 31 IleileAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArg 50
Db 97 ATCATCAACGAAGCCTTCTGGGTATTCATTTGCCGACCAACACCGATTGGTAACCTTGAGA 156
Qy 51 PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly 70
Db 157 TTCAGGACCCAGTTCATATCCCGGTTCTTGGATGGTCAAAAGTTCACTTCTTACGGT 216
Qy 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90
Db 217 CCATCTTGTATGCAACAAACCCGAGAGGTACCTACGAAGAAACTTTGCCAAGGCCAGCT 276
Qy 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110
Db 277 TTAGATCTGGTTATGCATCCAACT-----CCATCTTCTGAAGACTGT 320
Qy 111 LeuThrIleAsnValAlaArgProProGlyThrLysAlaGlyAlaAsnLeuProValMet 130
Db 321 TTGACCATTAATGTGTAGACACCCCGGCAAGGCTGGTGCCAACTTGCACAGTTATG 380
Qy 131 LeuThrIlePheGlyGlyPheGluIleGlySerProThrIlePheProAlaGln 150
Db 381 TTGGGATCTTTGGTGGGTTTGAAGTTGGTGCTAGTACTTCCCTCCAGCCCAA 440
Qy 151 MetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArg 170
Db 441 ATGATTACCAAGTCTATTGCTATGGTGAAGCCATCATCCAGCTTCTGTCACTACAGA 500
Qy 171 ValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAla 190
Db 501 GTCTCGAGCTGGGTTTCTTGCTGTGTCGCAAAATCAAGGCCGAAGGTTCTGCCAACGCC 560
Qy 191 GlyLeuLysAspGlnAtcLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly 210
Db 561 GGTGTGAAGGACCAAGGATGGGTATGCAATGGGTGGCTGACAACTTGGCTTTGGT 620
Qy 211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230
Db 621 GGTGATCCAACTAAGTTACTATCTTTGTGAATCTGCTGTTCTATGCTCGCTCATGTGT 680
Qy 231 HisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250
Db 681 CACATTTTGTGCAACGCGGTGACAACTTACAAGGGTAAGCCATTTGTCAGAGCTGGT 740
Qy 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270
Db 741 ATCATGCATCTGGTGTATGGTTCCATCTGACGCCGCTGACGGTATCATCGGTAAACGA 800
Qy 271 IleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290
Db 801 ATTTTGTACTTGTGGCTTCCAAACGCTGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 860
Qy 291 LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
Db 861 TTGAGAGGTGTTTCTTCTGACACTTTGGAAGACGCCCAACCAACACCCCTGGTTCTTG 920

```

```

Qy 311 AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330
Db 921 GCTTACTCCTCTTAAGATTGTTCTTACTTCCCAAGACCAAGCGGTGTAACATCACCGAC 980
Qy 331 AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350
Db 981 GACATGTACGCTTTGGTTAGAGAAGGTAAAGTAGTATGCCAACATCCTGTATTCATCGGTGAC 1040
Qy 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
Db 1041 CAAAACGACGAAGGTACCTTCTTGGTACTTCTTCTTGAACGTTACCATCTGATGCCAA 1100
Qy 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
Db 1101 GCCAGAGAAATATTTCAAGCAATCTTTGTCCCGTAGCGAGCTGAATTCGACACTTTG 1160
Qy 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
Db 1161 ATGACTGCTTACCCAGGTGACATCAAGGTTCCTCATTTGACACTGGAATTCATAAC 1220
Qy 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
Db 1221 GCCTTGACCCCAATTCAGAGAAATCTCTGCTGTTTGGTGACTTGGGTTTACTTTG 1280
Qy 431 AlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLys 450
Db 1281 GCTCGTAGATATCTTTGAACCACTACACCGGTGGTACCAAGTACTCTTCTGCTAAG 1340
Qy 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
Db 1341 CAATTGCTGCTGGTTGGCAGTTTGGGTACTTTCCACTCAACGATATCGTCTTCCAGAC 1400
Qy 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
Db 1401 TACTTGTGGTCTGGTCTGCTGATGATGATCAACGCTTTCATGCTTTGCTTCCACTGAC 1460
Qy 491 LeuAspProAsnThrAlaGlyLeuValLeuValAsnTrpProLysTyrThrSerSerSerGly 510
Db 1461 TTGGACCCAAACACCGCGGT-----TTCTTCTTCTCA 1493
Qy 510 nSerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPh 530
Db 1494 ATCTGGTAACAACTTGATGATGATCAACGCTTGGGTTTGTACACCGGTAAAGCAACTT 1553
Qy 530 eAtqThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db 1554 CAGAACCGCGGTACGCGCTTGTCTCCAACCCCATCTTCTTCTTGT 1605

RESULT 9
AX709921 1532 bp DNA linear PAT 10-APR-2003
LOCUS Sequence 3 from Patent EP1288294.
DEFINITION AX709921
ACCESSION AX709921.1 GI:29786302
VERSION AX709921.1
KEYWORDS Candida rugosa
SOURCE Candida rugosa
ORGANISM Candida rugosa
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
1
AUTHORS Tang, S.J., Lee, G.C. and Shaw, J.F.
TITLE Recombinant candida rugosa lipases
JOURNAL Patent: BP 1288294-A 3 05-MAR-2003;
Academia Sinica (TW)
FEATURES
source location/Qualifiers
1..1532
/organism="Candida rugosa"
/mol_type="unassigned DNA"
/db_xref="taxon:5481"

ORIGIN
Alignment Scores:

```

Pred. No.: 1,93e-191 Length: 1532  
 Score: 2392.50 Matches: 502  
 Percent Similarity: 94.37% Conservatives: 1  
 Best Local Similarity: 94.18% Mismatches: 9  
 Query Match: 83.54% Indels: 26  
 DB: 6 Gaps: 16

US-09-943-857-4 (1-547) x AX709921 (1-1532)

QY 15 ProThAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleLeuAsnGlu 34  
 DB 2 CCCACGCCCAAGCTGCCAAGCGGACACCATCACCGGTCTCAACGCCCATCATCAACGAG 61  
 QY 35 AlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro 54  
 DB 62 GCGTTCCTCGGCATTCCTTTGCGAGCGCGCGTGGCAACCTCCGCTTCAAGGACCT 121  
 QY 55 ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMet 74  
 DB 122 GTGCCGTACTTGGCTCGCTCAACGCCCAAGATT-ACT---TACGGCCG---TGCATG 174  
 QY 75 GlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuVal 94  
 DB 175 CAGCAGAACCCCGAGGCGCAGTTTGAGAGAACCTTGGCAAGCGGACCTCGACCTTGGTG 234  
 QY 95 MetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsn 114  
 DB 235 ATGCAGTCAAGGTGTTCAGGCGGTGCTTCCCGAGAGTGAGGACTGCCTCACCATCAAC 294  
 QY 115 ValValArgProProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuThrIlePhe 134  
 DB 295 GTGGTGGCGCGCGGCGCACCAAGGCGGCGCCCAACCTCCCGGTCTGCTCGATCTTT 354  
 QY 135 GlyGlyGlyPheGluIleGlySerProThrIlePheProProAlaGlnMetValThrLys 154  
 DB 355 GCGGTTGGGTTTGAGATCGGACGCCACCATCTTCCCTCCCGCCAGATGGTCAACAG 414  
 QY 155 SerValLeuMetGlyLysHisIleLeuHisValAlaValAsnTyrArgValAlaSerTyr 174  
 DB 415 AGTGTCTCATGGCAAGCC-ATCATCCAGTGGCGGTCACTACCGTGTGCTCGTGG 473  
 QY 175 GlyPheLeuAlaGlyAspAspIleLysAlaGlyLysSerGlyAsnAlaGlyLeuLysAsp 194  
 DB 474 GGGTCTTGGCTGTGATGACATCAAGCGCGAGCGGAGCGGCAACCGCGCTTGAAGGAC 533  
 QY 195 GlnArgLeuGlyMetGlnTyrValAlaAspAsnIleAlaGlyPheGlyGlyAspProSer 214  
 DB 534 CAGCGTTTGGGTCATGAGTGGTGGGAGACACATTTGCCGGTTCCGGCGGACCCGAGC 593  
 QY 215 LysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTyr 234  
 DB 594 AAGGTGACATC-TTTGGCGAG--GCGGCGAGCATGTCCGTGTGTGCGCACCTCATCTGG 649  
 QY 235 AsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSer 254  
 DB 650 AACGACGGCGCAACACGTACAAGGCGAGCGGTGTTCCGCGCGGCATCATGCGAG-- 706  
 QY 255 GlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeu 274  
 DB 707 GGAGCATGTGTCG--GACCGGTGGAGCGGACGATGCGGACGAGATCTACGACCTC 763  
 QY 275 PheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAla 294  
 DB 764 TTTGTCTCGAGTGTGGTGTGGCAGCGCCAGCAGCATCGCGTGTGCGCAGTGG 823  
 QY 295 SerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSer 314  
 DB 824 ---AGCGACACCTTGTGTCGATGCGACCAACACACTCTCTGGTCTTGGCGTACTCTCG 880  
 QY 315 LeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLys 334  
 DB 881 TTGCGGTG-----TACTCCGCGCGCGGCGGCAAGACATCACCATGATGTACAAG 934

QY 335 LeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAspGlnAsnAspGlu 354  
 DB 935 TTGGTGGCGGACGGCAAGTATGCAAGCGTTCCTCGTGTATCATTTGGCGACCAAGACGAG 994  
 QY 355 GlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyr 374  
 DB 995 GGCACCATCTTTGGCTC-----TTGAACGTGACCAAGTATGCTCAGGCCCGTGTCTAC 1047  
 QY 375 PheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaIaTyr 394  
 DB 1048 TTCAAGCAG---TTCATCCACGCCGAGCGAGATCGACACCTTGATGGCGGGGTAC 1104  
 QY 395 ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrPro 414  
 DB 1105 CCCCAGACATCACCCAGGGT---CCGTTCCGACAC---GGT-----TCAACGCTCACCCCG 1154  
 QY 415 GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgTyr 434  
 DB 1155 CAGTTCAAGAGATC---GCGGTGCTCGGCGACCTTGTCATTCATCCACGCCCGCGCTAC 1211  
 QY 435 PheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGly 454  
 DB 1212 TTCTCAACCACTTCCAGGGCGGCACCAAGTACTCGTTCTC---AAGCAGCTC---GGG 1265  
 QY 455 LeuProIleMetGlyThrPheHisAlaAsnAspIleValTyrGlnAspTyrLeuLeuGly 474  
 DB 1266 TTGCCAATCATGGGACCTTCCATGCCACGACATTTGTGGCAGGACTACTTGTGGGA 1325  
 QY 475 SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsn 494  
 DB 1326 AGCGGAGCGGTCTATCTACCAACACGCGTTTATCGCTTCCGCCACCGACTTGGACCCCAAC 1385  
 QY 495 ThrAlaGlyLeuLeuValAsnTyrProLysTyrThrSerSerGlnSerGlyAsnAsn 514  
 DB 1386 ACCCGGGGTGTGTGTGAAGTGGCCCAAGTACACGAGCAGC-----CAGGGCACAAC 1439  
 QY 515 LeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGly 534  
 DB 1440 TTGATGATCATCAACGCTTGGCTTGTACCGGCAAGGACAACTTCGACCGGTGGC 1499  
 QY 535 TyrAspAlaLeuMetThrAsnProSerPhePheVal 547  
 DB 1500 TAGCAGCGTGTATGATCAACCAACCG-----TCTTTGTG 1532

RESULT 10  
 LOCUS CCLIP5 1812 bp DNA linear PLN 30-JUN-1993  
 DEFINITION C.cylindracea LIP5 gene.  
 ACCESSION X66008.S55942  
 VERSION X66008.1 GI:296937  
 KEYWORDS lip5 gene; lipase.  
 SOURCE Candida cylindracea  
 ORGANISM Candida cylindracea  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 REFERENCE 1 (bases 1 to 1812)  
 AUTHORS Lotti,M., Grandori,R., Fusetti,F., Longhi,S., Brocca,S.,  
 Tramonano,A. and Alberghina,L.  
 TITLE Cloning and analysis of Candida cylindracea lipase sequences  
 JOURNAL Gene 124 (1), 45-55 (1993)  
 MEDLINE 93178975  
 PUBMED 840480  
 REFERENCE 2 (bases 1 to 1812)  
 AUTHORS Longhi,S.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-MAY-1992) S. Longhi, Universita' degli Studi di  
 Milano, Via Celoria 26, 20133 Milano, ITALY  
 COMMENT See X66006-8, X16712, X64703 & X64704  
 X66006, X66007 and X66008 are related sequences in the description  
 of C.cylindracea Lipi and Lip2 sequences (x64703, x64704).  
 FEATURES  
 Location/Qualifiers  
 1..1812  
 /organism="Candida cylindracea"



```

/mol_type="genomic DNA"
/strain="ATCC 14830"
/db_xref="taxon:4322"
/clone_lib="SACI genomic"
116..1765
/gene="LIP5"
116..1765
/gene="LIP5"
/EC_number="3.1.1.3"
/codon_start=1
/product="triacylglycerol lipase"
/protein_id="CAA46807.1"
/db_xref="GI:1325991"
/db_xref="COA:P32949"
/db_xref="SWISS-PROT:P32949"
/translation="MKLALALLIASVAAPATATLANGDTTIGLNALINEAFGLIPFA
EPPVGNLRFDPVPRVPSLNGQSFATYGLPQMOPREGTYEENLPKVALDLVMSQKVP
QAVLPNSBCLTINVPRPGTKAGANLPVLMWIFGGGFEIGSPTIFPPAOMVSKSLM
KPTLHVAVNYRLASFGLAGPDIKAGSSNAGLKQDRLGMQWADVADNAGFGDFSKV
TIFELAGSSVLCHLLWNGDNTYKQKPLFRAGIMOLGAMVPLDPVDGTGTGTIDT
LVALTGSSASNKLAICLRLTLQALLDATNDTPGFLSVTSRLILYLRPPDGNATIDDM
YKVRDGYASVPVLIQDNDGEGELFGLLLNTTTEADAEAYLRKLFIHATDADITAL
KAAVPSVDVTQGLPDTGILNALTQPKRINAVLGLTFTLSRRYFLNHYTGGPKYSFL
LKQLLPLIGTFHNDIVQHFLLGSGSVIYNNAFTAFATDLDLPNTAGLLVQWPKST
SSLQAGDLMQISALGLYTGKONFRTAGYNALFADPLHFFV"
116..250
/gene="LIP5"
251..1762
/gene="LIP5"
/product="triacylglycerol lipase"
/EC_number="3.1.1.3"

sig_peptide
mat_peptide

ORIGIN
Alignment Scores:
Pred. No.:      1.61e-187      Length:      1812
Score:          2347.00      Matches:      445
Percent Similarity: 88.83%      Conservative: 32
Best Local Similarity: 82.87%      Mismatches:   60
Query Match:      81.95%      Indels:       0
DB:               8           Gaps:         0

US-09-943-857-4 (1-547) x CCLIP5 (1-1812)

Qy 11 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30
Db : : : : :
Db 152 GTGGTGTGTCGCCCCCACCAGCGCGCTCGCCAAACGGCGACACCATCACCGGCTCAACGCC 211
Qy 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArg 50
Db : : : : :
Db 212 ATCATCAACGAGCGGTCTCTCGGCATTCCTTTGCGAGCGCGCGGTGGGCAACCTCCGC 271
Qy 51 PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly 70
Db : : : : :
Db 272 TTCAAGGACCCCTGTGCGGTACCGTGGTCTCTCAACGGTCAATCCTTCACCGCGTACCGT 331
Qy 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90
Db : : : : :
Db 332 CGCTGTGTCATGACGAGAACCCCGAGGCGACCTACGAGGAGAACTCCCCAAGGTGGCG 391
Qy 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110
Db : : : : :
Db 392 CTTGACTTGTGATGCAGTCCAAAGGTGTTCCAGGCTGTTCTCCCCAACAGCGAGACTGC 451
Qy 111 LeuThrIleAsnValValArgProProGlyThrLysAlaGlyAlaAsnLeuProValMet 130
Db : : : : :
Db 452 CTCACCATCAACGTGTGCGCGCGCGGACCAAGGCGGCGGCAACCTCCCGGTGATG 511
Qy 131 LeuTyrIlePheGlyGlyGlyPheGluIleGlySerProThrIlePheProAlaGln 150
Db : : : : :
Db 512 CTTGTGATCTTTGGCGGTGGGTGTGATCGGACGCCAGCCACCATCTCTCCCTCCCGTCA 571
Qy 151 MetValThrLysSerValLeuMetGlyLysIleIleHisValAlaValAsnTyrArg 170
Db : : : : :

```

```

Db 572 ATGGTCTCCAAGAGTGTGCTCATGGGCAAGCCCAATCATCCACGTGGCGGTCAATCACC 631
Qy 171 ValAlaSerTrpGlyPheLeuAlaGlyAspIleLysAlaGluGlySerGlyAsnAla 190
Db : : : : :
Db 632 TTGGCGTCTTTGGTTTCTTGGCGGTCCGGACATCAAGGCGGAGGAGGTCCAAATGCC 691
Qy 191 GlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly 210
Db : : : : :
Db 692 GGCCTCAAGGACCAAGCGCTTGGGCATGCAGTGGGTGGCAGACAACATTCGCGGTTCGCG 751
Qy 211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230
Db : : : : :
Db 752 GGCACCCGAGCAAGGTGACCATCTTTGGTGAGTGGCGGACGATGTCTCGTGTGTGTC 811
Qy 231 HisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250
Db : : : : :
Db 812 CACCTTCTCTGGAATGGCGGCGACACACGATCAAGGCGCAACGCTGTCTTCGCGCGGCG 871
Qy 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270
Db : : : : :
Db 872 ATCATGCAGCTGGGAGCCATGCTGCCGCTGGACCCGCTGGACGCGCATATGGAAACCCAA 931
Qy 271 IleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290
Db : : : : :
Db 932 ATCTATGACACGTGTGGTGGCTCTGACGCGGCTGCAGAGTGCAGCAACAGCTTCGCTGC 991
Qy 291 LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
Db : : : : :
Db 992 TTGGTGTGCTCTGACTCAGGCATTTGCTCATGCCCAACAGCACCCCTGGGTCTTGTG 1051
Qy 311 AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330
Db : : : : :
Db 1052 TCGTACACCTGTTGGGTGCTGCTGTACTCTCCCGCGCGCGGCGCAACATCACCGCAT 1111
Qy 331 AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350
Db : : : : :
Db 1112 GACATGTACAAGTGTGTACGCGAGCGCAAGTATGCAACGCTTCCGCGTATCATTTGGCGAC 1171
Qy 351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370
Db : : : : :
Db 1172 CAGAACGACGAGGGCTTCTGTTGGTCTCTCGTGTGTTGAACACCCACCGAGCGCGAC 1231
Qy 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
Db : : : : :
Db 1232 GCGGAGGCATACCTCAGAAAGCTGTTCATCCACGCCACCGCGCGCATATCACCGCATTG 1291
Qy 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
Db : : : : :
Db 1292 AAGCGCGGTACCCCGAGCATGTACCCAGGGGTCTGCGGTTCGACGCGGCGCATTTCAAC 1351
Qy 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
Db : : : : :
Db 1352 GCCCTTACACCCAGCTCAAGCGGATCAATGCTGTGCTTGGCGACCTCACCTTACCCTC 1411
Qy 431 AlaArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLys 450
Db : : : : :
Db 1412 TCGCGCGCTACTCTCTCAACCACTACACCGGTGTGCCAGTACTCGTTCCTCTCTGAAG 1471
Qy 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
Db : : : : :
Db 1472 CAGCTTCTGGGCTTCCCATCTCTCGGCAGCTTCCACGCGGTTCATCGCGTTTCCACCGAC 1531
Qy 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
Db : : : : :
Db 1532 TTTTGTGGGCGAGCGGCGAGCGTCTATCAACAACGCGTTCATCGCGTTTCCACCGAC 1591
Qy 491 LeuAspProAsnThrAlaGlyLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db : : : : :
Db 1592 TTGGACCCCAACACCGCGGCTTGTGTGTGAGTGGCGGCAAGTCCACGAGCGCTGCGAG 1651
Qy 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
Db : : : : :
Db 1652 GCGGGGACCAACTTGTATGCAGATCATGCTTGGGCTGTATACCGCGCAAGGACCACTTC 1711

```





```

Db 1155 CAGAACACGAGGCGACAGGTTTGGCTGCTGTGTGAACGTAACGATGCTCAG 1214
Qy 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrIleu 390
Db 1215 GCACGCGAGTACTTCAAGGAAGTCTCATCCACGCCAGCGAGATCGACACCTTG 1274
Qy 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValIleuAsn 410
Db 1275 ATGCGCGGTACCCAGCGACATCACCCAGGGTCTGCCGTTTCGACACCGCATCTTCAAC 1334
Qy 411 AlaLeuThrProGlnPheLysArgIleSerAlaValIleuGlyAspLeuAlaPheIleHis 430
Db 1335 GCCATCACCCGCGATTCACACGAGTTGCAGCGGTGCTTGGTACCTTGGTCACTCTC 1394
Qy 431 AlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheIleuSerLys 450
Db 1395 CCCCGGCGTACTTCCCTCAACCACTTCCAGGCGCGCACCAAGTACTCGTCTCCTGAAG 1454
Qy 451 GlnIleuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrioGlnAsp 470
Db 1455 CAGCTTCTGGGTTTCCCGGTGATTCGACCCACCCACCGACGACATGTTGTGGCAGAC 1514
Qy 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
Db 1515 TTTTGGTGAGCCACGACGCGCGCTGTACACACGCGTTTATTGCTTTGCCAACGAC 1574
Qy 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTyrProLysTyrThrSerSerGln 510
Db 1575 CTCGACCCGAAACAAGCGCGTTTGTGTGAACCTGGCCCAAGTACACCGACGCGTGCAG 1634
Qy 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
Db 1635 CTGGGCAACAACCTTGTGAGATCAACCGCTTGGCTGTATACCGCGGACGACAACTTC 1694
Qy 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db 1695 CGACCGCTGGCTACGACGCGTTTGTGTGAACCTGGCCCAAGTACACCGACGCGTGCAG 1745

RESULT 12
LOCUS CCLIP2 2043 bp DNA linear PLN 27-APR-1993
DEFINITION C.cylindracea LIP2 gene for lipase.
ACCESSION X64704
VERSION X64704.1 GI:2546
KEYWORDS Candida cylindracea
SOURCE Candida cylindracea
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 2043)
AUTHORS Longhi,S., Fusetti,F., Grandori,R., Lotti,M., Vanoni,M. and
Alberghina,L.
TITLE Cloning and nucleotide sequences of two lipase genes from Candida
cylindracea
JOURNAL Biochim. Biophys. Acta 1131 (2), 227-232 (1992)
MEDLINE 92305068
PubMed 1610906
REFERENCE 2 (bases 1 to 2043)
AUTHORS Longhi,S.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-1992) S. Longhi, Universita' degli Studi di
Milano, Via Celoria 26, 20133 Milano, ITALY
COMMENT See also X64703
X66006, X66007 and X66008 are related sequences in the description
of C.cylindracea Lip1 and Lip2 sequences (X64703, X64704).

FEATURES
Location/Qualifiers
1..2043
/organism="Candida cylindracea"
/mol_type="genomic DNA"
/strain="ATCC 14830"
/db_xref="taxon:44322"
/clone="B"

```

```

TATA_signal 126..132
gene 161..1807
/ gene="LIP2"
CDS 161..1807
/ gene="LIP2"
/ product="triacylglycerol lipase"
/ EC_number="3.1.1.3"
/ codon_start=1
/ protein_id="CAA45958.1"
/ db_xref="GI:1325988"
/ db_xref="GOA:P32946"
/ translation="MKLUCLALGAARAAAPTATLANGDTITGLNALVNEKELGIPPEAE
PVVTLEKFPVPYSASINGQQFTLYGLPCQMNPMSFEDTLFNNARHLVLQSKIFQ
VVLPNDSGLCTINIRPPFTRASAGLFLMWIFGGFELGSSLPFGDQMVAKSVLMG
KPVTHGSMYRVASWGLAGPDIQNEGSAGLDORLAMQVADNATAGFGDPSKVT
IYGLAGSMSTFVHLVWNGDNTYKGPLPRAAIWOLGCMVLPDVPDGTGTETYNQV
VASGCSASDKLACLGLQDLYQATSDTPGVLAVERSLRLLVLPDPDGTFTIDDMY
ALVRDGYAHVFIIGQNDDEGLTGLLLLVTTDAQARVPKOLFTHASDAEIDTLM
AAVTSIDITQGLPFDTFINAIPTQKRILALLGLDIAFLARRYFLNYQGGTKYSELL
KQLGLFLVLTGFHNDIITWQDYLVGSGSVIYNNAFIAPANDLDPNKAGLWNTWFTYS
SLQGNLMLQINGLGLYTGKDNFRPDAYSALFSNPPLFFV"
161..202
/ gene="LIP2"
/ note="triacylglycerol lipase"
203..1804
/ gene="LIP2"
/ product="triacylglycerol lipase"
/ EC_number="3.1.1.3"

sig_peptide
mat_peptide

ORIGIN
Alignment Scores:
Pred. No.: 4,14e-176 Length: 2043
Score: 2212.00 Matches: 417
Percent Similarity: 84.17% Conservative: 35
Best Local Similarity: 77.65% Mismatches: 85
Query Match: 77.23% Indels: 0
DB: 8 Gaps: 0

US-09-943-857-4 (1-547) x CCLIP2 (1-2043)
Qy 11 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30
Db 194 GTGGCGGACGACCCGCCACGCGCACCTTCGCCAACGCGGACACCATCACCGCTCTCAACGCG 253
Qy 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArg 50
Db 254 ATTGTCAACGAAAGTTTCTCGGCATACCGTTTCCGAGCGCGCCGCGTGGCAGCTCCGC 313
Qy 51 PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly 70
Db 314 TTCAAGCGCGCGTCCGCTACTCGCGCTCGCTCAACGCGGACGAGTTTACCCTGTACGCG 373
Qy 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90
Db 374 CCGCTGTGATGACAGATGAACCTATGCGCTCGTTGAGGACACACTTCCCAAGATGCG 433
Qy 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110
Db 434 CGGCATTTTGGTCTCCAGTCCAAGATCTTCCAAGTGTGCTTCCCAACGACGAGGACTGT 493
Qy 111 LeuThrIleAsnValValArgProProGlyThrLysAlaGlyAlaAsnLeuProValMet 130
Db 494 CTCACCATCAAGTGATCCGCGCGCGCGGACACGAGCGCGAGTGTGGTCTCCCGGTGATG 553
Qy 131 LeuTyrIlePheGlyGlyGlyPheGluIleGlySerProThrIlePheProAlaGln 150
Db 554 CTCGTGAUTTTGGCGGTGGGTTTGAGCTTGGCGGCTCCAGCGCTCTTCCAGGAGACGAC 613
Qy 151 MetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArg 170
Db 614 ATGGTGGCCAAAGAGCGTGTCTCATGGGTAAACCGGTGATCCACGTGAGCATGAACACTCCGC 673

```

171 ValAlaSerTrpGlyPheLeuAlaGlyAspAspIleHisAlaGluGlySerGlyAsnAla 190  
Db GTGGCGTCATGGGGTCTTGGCGGCCCCGACATCCAGAACGAGGAGCGGAAACGCC 733

191 GlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly 210  
Db GTGGCGTCATGGGGTCTTGGCGGCCCCGACATCCAGAACGAGGAGCGGAAACGCC 733

211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230  
Db GGCAGCCCGAGCAGGTGACCATATACGGCGAGCTGGCGGCGACATGTCCAGCTTTGTG 853

231 HisIleuTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250  
Db CACCTTGTGTGGACGACGCGGCGACACACGCTACCAACGCGAAGCGGTTGTTCCGCGCGCC 913

251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270  
Db ATCTACACCGAGTGTGGCGTCTGCGGGTGTGGAGTGCACGACGACAGTCCGCTGC 1033

291 LeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310  
Db TTGCGCGCGCTTCTGCAGGACACGCTTACCGAGCCACGACGACGCGCGCGCTGTG 1093

311 AlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330  
Db GCGTACCGCGTGTGGCGTGTGCTATCTCCCGCGCGCGCGACGCTTCATCACCAGC 1153

331 AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350  
Db GACATGATGCTTGTGGCGGCGGAGCGGAGTACGACACGCTGCGGTGATCATCGGCGAC 1213

351 GlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGln 370  
Db CAGAACGACGAGGCGCATTTGTTGGGCTCTGCTGTGTTGAACGTGACACAGATGCTCAG 1273

371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390  
Db GCACGGCGGTACTTCAAGCAGCTGTTCATCAGCGCGGACGATCGGAGATCGACACGTTG 1333

391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410  
Db ATGGCGCGGTACACCGACATCACCGAGGCTGCGCTTGCACCGCGCATCTTCAAT 1393

411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430  
Db GCCATCACCGCGAGTTCAAACGATCTCTGGCGTGTCTTGGCGACCTTGGCGTTACGCTT 1453

431 AlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLys 450  
Db GCGCGTCTACTTCTCAACTACTACCGGCGGCGACCAAGTACTGTTCTTCCTGAAG 1513

451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470  
Db CAGCTTCTGGGTTGCGGCTTGGGACCTTCCACGGCGACGACATCATCTGGCAGGAC 1573

471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaThrAsp 490  
Db TACTTGTGGCGCGCGCAGTGTGATCTACAACACGCGTTCATTGGGTTTGGCAACGAC 1633

491 LeuAspProAsnThrAlaGlyLeuValAsnTrpProLysTyrThrSerSerSerGln 510  
Db CTGACCCGGAACAGGGCGGCTTGTGGACCAACTTGGCGCCACGTACACGACGCTGTGCAG 1693

511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530  
Db CTGGCGAACAACTTGATGCAGATCAACGGCTTGGGGTGTGTACACCGCGCAGGACCACTTC 1753

531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547

1754 CGCCCGGATGGTACAGCGCCTCTTTTCCAAACCGCCACTGTTCTTTGTG 1804

RESULT 13  
AX709925  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Candida rugosa  
Candida rugosa  
Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE  
1  
Tang, S. J., Lee, G. C. and Shaw, J. F.  
Recombinant candida rugosa lipases  
Patent: EP 128294-A 7 05-MAR-2003;  
Academia Sinica (TW)

FEATURES  
Location/Qualifiers  
source  
1..1541  
/organism="Candida rugosa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:5481"

ORIGIN  
Alignment Scores:  
Pred. No.: 2,53e-161 Length: 1541  
Score: 2034.00 Matches: 434  
Percent Similarity: 87.57% Conservative: 31  
Best Local Similarity: 81.73% Mismatches: 49  
Query Match: 71.02% Indels: 24  
DB: 6 Gaps: 15

US-09-943-857-4 (1-547) x AX709925 (1-1541)

QY 17 AlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleAsnGluAlaPhe 36  
Db 7 GCCACGCTCGCCAAACGGGACACCATCACCGGTCTCAACGCCATCATCAACGAGGCTTC 66

QY 37 LeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspProValPro 56  
Db CTGCGCATTCCTTTGCGGAGCGCGGTGGCAACCTCCGCTTCAAGGACCTGTGCG 126

QY 57 TyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGln 76  
Db TACGTGGGTCTCTCAACGGTCAATCTTCCCGCGTACGCTCCG---TGCATGCACGAG 183

QY 77 AsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGln 96  
Db AACCCCGAGGCGACCTACGAGGAGAGACCTCCCAAGGTGGCGCTTGACTTGGTGTGTCAG 243

QY 97 SerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValVal 116  
Db TCCAAGGTGTTCAGGTGTCTTCCCAACAGGAGGACTGCGCTCACCATCAACGCTGGTG 303

QY 117 ArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGly 136  
Db CGGCCCGCGGCGACCAAGCGGCGGCAACCTCCCGTCTATGCTCTGGATCTTTTGGCGGT 363

QY 137 GlyPheGluIleGlySerProThrIlePheProAlaGlnMetValThrLysSerVal 156  
Db GGGTTTGTAGATCGGACGCGCCACCATCTTCCCTCCCGCTCAGATGCTCCAGAGTGTG 423

QY 157 LeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrpGlyPhe 176  
Db CTATCGGCGAGCC-ATCATCCAGTGGCGGTCAACTACCGCTTGGCTCTTGTGTTTC 482

QY 177 LeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArg 196  
Db TTGGCGGTCGCGACATCAAGCGGCGGCGAGCTCCCAATGCGCGGCTCAAGGACGAGCGC 542

QY 197 LeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysVal 216

Db	543	TTGGGATGAGTGGTGGCAGACACATTCGCGGGTTCGGCGGACCCGAGCAAGTG	602		AX709923	1548 bp	DNA	linear	PAT 10-APR-2003
Qy	217	ThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAsp	236		SEQUENCE	5 from Patent EP1288294.			
Db	603	ACCATCTTTGGGAG-----GCGGGCAGCATGTCCGTGTGTGGCCACCTTCTCTGGAATGGC	658		AX709923				
Qy	237	GlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAla	256		AX709923.1	GI:29786303			
Db	659	GGCGACACACGTACAAGGCAAGCGTGTTCGCGGGCATCATGACG---GGAGCC	715		KEYWORDS	Candida rugosa			
Qy	257	MetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheVal	276		ORGANISM	Candida rugosa			
Db	716	ATGGTGCGG---GACCCCGTGGACGGCACTATGGACC-CAAAATCATGACAGTTGGTG	771		REFERENCE	Tang, S.J., Lee, G.C. and Shaw, J.F.			
Qy	277	SerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSer	296		AUTHORS	Recombinant candida rugosa lipases			
Db	772	GCTACG---GGCTGAGCAGTGCACCAACAAAGCTTGGCTGCTTGGTGGTCTT---ACT	825		TITLE	Patent: EP 1288294-A 5 05-MAR-2003;			
Qy	297	AspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArg	316		JOURNAL	Academia Sinica (TW)			
Db	826	CAGGCATTGCTCGATGCCACCAACGACACCCCTGGGTTCTTGTCGTACACCTCGTTGGG	885		FEATURES	Location/Qualifiers			
Qy	317	LeuSerTyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLysLeuVal	336		1. 1548				
Db	886	TTG-----TACTCCGCGCCGACGCGCCACATCCGATGACATGTACAAAGTTGGTA	939		source	/organism="Candida rugosa"			
Qy	337	ArgAspGlyLysTyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyThr	356		/mol_type="unassigned DNA"				
Db	940	CSCGACGGCAAGTATGCAAGCGTTCCCGTGTATCAVTGGCGACCAAGACGAGGGCTTC	999		/db_xref="taxon:5481"				
Qy	357	IlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLys	376		ORIGIN				
Db	1000	TTGTTTGT-CTC-----TTGAACACCACCGAGCGCGGACGATACCTCAGA	1052		Alignment Scores:				
Qy	377	GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGln	396		Pred. No.:	3.82e-160	Length:	1548	
Db	1053	AAG---TTCATCCACCGCACCGCGATATCACCGCATTCAGCGCGGTACCCACGC	1109		Score:	2020.00	Matches:	422	
Qy	397	AspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPhe	416		Percent Similarity:	87.12%	Conservative:	38	
Db	1110	GATGTCAACCCAGGGT---CCGTTCGACACGGGCATTTCTCAACGCCCTTACACCCAGCTC	1166		Best Local Similarity:	79.92%	Mismatches:	57	
Qy	417	LysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeu	436		Query Match:	70.53%	Indels:	17	
Db	1167	AAGCGATCAATGCTGTGCTTGGCGACCTCACTTTACCTTCGCGCGCTACTTCCTC	1226		DB:	6	Gaps:	9	
Qy	437	AsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuPro	456		US-09-943-857-4 (1-547) x AX709923 (1-1548)				
Db	1227	AACCACTACACCGGTGTCCTCAAGTACTCGTTCTCTC---AAGCAGCTT---GGTTG-CC	1279		Qy	15	ProthrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGlu	34	
Qy	457	IleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGly	476		Db	2	CCCCTGCCACGCTCGCCAAACGGCGACACCATCACCGGTCTCAACGCCATCATCAACGAG	61	
Db	1280	ATTCTCGGACG-TTCCACCGCAACGACATTTGTGGCAGACATTTTGTGGCGACGCGC	1338		Qy	35	AlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspPro	54	
Qy	477	SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla	496		Db	62	CGCTTCCCTCGGTATTTCCTTGTCTCAGCGCGCGGTGGGCACTCCGCTTCAAGCGCGCT	121	
Db	1339	ACGCTCATCTACAACACGCTTCATCGGTTTGGCACCGACTTGGACCCCAACACCGCG	1398		Qy	55	ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMet	74	
Qy	497	GlyLeuLeuValAsnTrpProLysTyrThrSerSerGlnSerSerGlyAsnAsnLeuMet	516		Db	122	GTGCGCGTACTCGGCGTCTCTCAATGGTCAGAAGTTACTGTATGG---CCCTG-TCGATG	177	
Db	1399	GGC---TTGGTGCAGTGGCCCAAGTC-ACC---AGCAGCGAGCGCGGAGACTGTCTCACC	1451		Qy	75	GlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuVal	94	
Qy	517	MetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAsp	536		Db	178	CAGATGAACCCATTGGGCAACTGGGACTCTCGCTTCCCAAGGCTGCCATCAACTTG---	234	
Db	1452	CAGATCAGTGCCTTGGGCTTGTACACCGCAGGAGCACTTCGCGACCCCGGCTCAAC	1511		Qy	95	MetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsn	114	
Qy	537	AlaLeuMetThrAsnProSerSerPhePheVal	547		Db	235	ATGCAGTCCAGCTCTTCAGCGGTGTCTTAAACGGCGAGGACTGTCTCACCATCAAC	294	
Db	1512	GCTTTGTTGGCGACCCG---CACTTTTTCGTG	1541		Qy	115	ValValArgProProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePhe	134	
RESULT 14					Db	295	GTGHTGCGG---CCGGGACCAAGCGGGGTGCCAACCTCCCGGTGATGTGTGGAITTTT	351	
AX709923					Qy	135	GlyGlyGlyPheGluLeuGlySerProThrIlePheProProAlaGlnMetValThrLys	154	
					Db	352	GGCGCGGGTTTGAGGTGGCGGCTCCAGTCTCTTCCCTCCCGCAGACAGATGATCACCGC	411	
					Qy	155	SerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrp	174	
					Db	412	AGCGTGTCTTATGGGCAAGCCCATCTCCACGTGAGCATGAATACCGGTTTGTTCGTGG	471	
					Qy	175	GlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAsp	194	
					Db	472	GGGTCTTGGCTGGTCCAGACATCAAGGCGGAGGCGAGGAGCCCGGTTTGACACGAC	531	
					Qy	195	GlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyAspProSer	214	
					Db	532	CAACGCTTGGGTGTGAGTGGGTGGCGGACAACTTGGCGGGTTCGGCGGAGCCCGTCC	591	
					Qy	215	LysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrp	234	

Db 592 AAGGTGACCATCTTTGGTGAGG--CGGGGACGATGTCGGTAAGTGTACGCTCTCTGG 649  
 QY 235 AsnAspGlyAspAsnThrTyrTyrGlyLysProLeuPheArgAlaGlyIleMetGlnSer 254  
 Db 650 AACGACGCGGACACACGTCACACGCGAAGCGGTGTTCCGTGGCGCCATCATGCAG--- 706  
 QY 255 GlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeu 274  
 Db 707 GGGGCCCATGGTGGCGG--GACCCGGTGGATGGGCCCTACGGCAGCGCATCTAGCACCG 764  
 QY 275 PheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAla 294  
 Db 765 GTGGTTGCTTCAGCGCGCTGTGGAGTCCAGCAGCAAGCTCGGTGCTTCGGCA--GCA 822  
 QY 295 SerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSer 314  
 Db 823 TCGAACGACAACTCTTCCAGGCCACCGACGACACTCGGGGGCTTGGGTATCCCTCG 882  
 QY 315 LeuArgLeuSerTyrLeuProArgProAspGlyLysAlaIleThrAspAspMetTyrLys 334  
 Db 883 TTGGCGGTGG--TTTCTCCCGCGCGCCGACGCGACCTTTCATCAGCATGATGTTCAAG 940  
 QY 335 LeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAspGlnAsnAspGlu 354  
 Db 941 TTGGTGGCGGACGCAAGTGTCCACAGCTTCGGTGATCATTTGGCGACCAAGACGAG 1000  
 QY 355 GlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyr 374  
 Db 1001 GGCACAGTGTTCGGTTG-----TTGAACGTGACTACGGATGCTCAGGACGCGCATC 1054  
 QY 375 PheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr 394  
 Db 1055 TTCAAGGAA--TTCAATCCACGCGCAGCGAGATCGACACTTGTATGGCGGCGTAC 1111  
 QY 395 ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrPro 414  
 Db 1112 CCAGCAGCATATCCACAGGCT---CGGTCGACACCGCGCATCTTCAACGCGCATCCCG 1168  
 QY 415 GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyr 434  
 Db 1169 CAGTTCAAACGGATTCGACGGTGTCTTGGTGACCTTGGTTCCTCTCCCGCGCGCTAC 1228  
 QY 435 PheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGly 454  
 Db 1229 TTCTCTCAACCATCTTCAGGCGCGCACCAAGTACTCGTTCC--TCGAAGCAGCTT---GGG 1283  
 QY 455 LeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGly 474  
 Db 1284 TTGGCGGTGATGGCACCCACACGCGCATGTTGGTTCCTCTCTCCCGCGCGTAC 1343  
 QY 475 SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsn 494  
 Db 1344 CACAGCAGCGCGGTGTACAAACGCGGTTTATTGGCTTTTGCCAAACGACCTCGACCCGAC 1403  
 QY 495 ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsn 514  
 Db 1404 AAGCGCGGTGTGTTGTGAAGTGGCCCAAGTATACACGAGCAGC-----CAGGGCAAC 1457  
 QY 515 LeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGly 534  
 Db 1458 TTGTTGCAGTCAACGCTTGGGCTGTATACCGCGCAAGCAACTTCGCGACCGCTGGC 1517  
 QY 535 TyrAspAlaLeuMetThrAsnPro 542  
 Db 1518 TAGCAGCGGTGTTTACCAACCG 1541

## RESULT 15

AX709927

LOCUS

DEFINITION

Sequence 9 from Patent EP1288294.

ACCESSION

AX709927

VERSION

AX709927.1

KEYWORDS

GI:29786305

## SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-943-857-4 (1-547) x AX709927 (1-1511)

QY 17

Db 7

QY 37

Db 67

QY 57

Db 127

QY 77

Db 180

QY 97

Db 240

QY 117

Db 294

QY 137

Db 354

QY 157

Db 414

QY 177

Db 474

QY 197

Db 534

QY 217

Db 593

QY 237

Db 650

QY 17

Db 7

QY 37

Db 67

QY 57

Db 127

QY 77

Db 180

QY 97

Db 240

QY 117

Db 294

QY 137

Db 354

QY 157

Db 414

QY 177

Db 474

QY 197

Db 534

QY 217

Db 593

QY 237

Db 650

QY 17

Db 7

QY 37

Db 67

QY 57

Db 127

QY 77

Db 180

QY 97

Db 240

QY 117

Db 294

QY 137

Db 354

QY 157

Db 414

QY 177

Db 474

QY 197

Db 534

QY 217

Db 593

QY 237

Db 650

QY 17

Db 7

QY 37

Db 67

QY 57

Db 127

QY 77

Db 180

QY 97

Db 240

QY 117

Db 294

QY 137

Db 354

QY 157

Db 414

QY 177

Db 474

QY 197

Db 534

QY 217

Db 593

QY 237

Db 650

QY 17

Db 7

QY 37

Db 67

QY 57

Db 127

QY 77

Db 180

QY 97

Db 240

QY 117

Db 294

QY 137

Db 354

QY 157

Db 414

QY 177

Db 474

QY 197

Db 534

QY 217

Db 593

QY 237

Db 650

QY 17

Db 7

QY 37

Db 67

QY 57

Db 127

QY 77

Db 180

QY 97

Db 240

QY 117

Db 294

QY 137

Db 354

QY 157

Db 414

QY 177

Db 474

QY 197

Db 534

QY 217

Db 593

QY 237

Db 650

QY 17

Db 7

QY 37

Db 67

QY 57

Db 127

QY 77

Db 180

QY 97

Db 240

QY 117

Db 294

QY 137

Db 354

QY 157

Db 414

QY 177

Db 474

QY 197

Db 534

QY 217

Db 593

QY 237

Db 650

QY 17

Db 7

QY 37

Db 67

QY 57

Db 127

QY 77

Db 180

QY 97

Db 240

QY 117

Db 294

QY 137

Db 354

QY 157

Db 414

QY 177

Db 474

QY 197

Db 534

QY 217

Db 593

QY 237

Db 650

QY 17

Db 7

QY 37

Db 67

QY 57

Db 127

QY 77

Db 180

QY 97

Db 240

QY 117

Db 294

QY 137

Db 354

QY 157

Db 414

QY 177

Db 474

QY 197

Db 534

QY 217

Db 593

QY 237

Db 650

QY 17

Db 7

QY 37

Db 67

QY 57

Db 127

QY 77

Db 180

QY 97

Db 240

QY 117

Db 294

QY 137

Db 354

QY 157

Db 414

QY 177

Db 474

QY 197

Db 534

QY 217

Db 593

QY 237

Db 650

QY 17

Db 7

QY 37

Db 67

QY 57

Db 127

QY 77

Db 180

QY 97

Db 240

QY 117

Db 294

QY 137

Db 354

QY 157

Db 414

QY 177

Db 474

QY 197

Db 534

QY 217

Db 593

QY 237

Db 650

QY 17

Db 7

QY 37

Db 67

QY 57

Db 127

QY 77

Fri Aug 6 10:53:11 2004

us-09-943-857-4.rge

```

QY 257 MetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheVal 276
Db |||||
QY 707 ATGGTC---CGGACGC-GTGGACGGCTC-TACGGCAACGAGATCTTTGACCTCTGGCG 761
Db |||||
QY 277 SerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSer 296
Db |||||
QY 762 TCGAC-GCGGGCTCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 817
Db |||||
QY 297 AspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArg 316
Db |||||
QY 818 GACACGTTGGAGGCGGCCACCAACACACCCCTGGGTTCTTGGCGTACTCTCTCGTTGGG 877
Db |||||
QY 317 Leu-SerTyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLysLeuVa 336
Db |||||
QY 878 TTGTACT-----CCCCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 928
Db |||||
QY 336 LArgAspGlyLysTyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyTh 356
Db |||||
QY 929 TCGGAGGGGCAAGTATGACT-----CCTGTGATCATCGCGCACGACGACGAGGGCAC 984
Db |||||
QY 356 rIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLy 376
Db |||||
QY 985 CTCTCTTTGGC-----ACCTTGACGTCACACCGAT-----GCCAGCGCGCGAGTACT 1032
Db |||||
QY 376 sGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGl 396
Db |||||
QY 1033 TCAGCAGTTTGTCTCCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1089
Db |||||
QY 396 nAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPh 416
Db |||||
QY 1090 -GACATACCCA-GGG---CCGTTTCGACACGGGTTT-CTCAACGCGCTCACCGCGAGTT 1143
Db |||||
QY 416 eLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyrPheLe 436
Db |||||
QY 1144 CAAGAGATC---GCGGTCTCGGCGACCTTGCTTACGC-----CGTCGCTACTTCT 1193
Db |||||
QY 436 uAsnHisPheGlnGlyThrIlyThrIlySerPheLeuSerLysGlnLeuSerGlyLeuPr 456
Db |||||
QY 1194 CAACCACTACACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1247
Db |||||
QY 456 oIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuGlySerGl 476
Db |||||
QY 1248 GGTGCTCGGACGTTCCACTCCACGACATGTCTCCAGGACTACTTGTGGCGAGCGG 1307
Db |||||
QY 476 ySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAl 496
Db |||||
QY 1308 CTGCTCATCTACAACAACGCGTTTCATTGCGTTGCGGCGGCGGCGGCGGCGGCGGCGG 1367
Db |||||
QY 496 aGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMe 516
Db |||||
QY 1368 GGGGTTGTTGGTGAAGTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1421
Db |||||
QY 516 tMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAs 536
Db |||||
QY 1422 GATGATCAACGCTTGGGCTTGTACACCGGCAAGGACAACTC-CGCACGCGCGGCTACGA 1480
Db |||||
QY 536 pAlaLeuMetThrAsnProSerSerPhePheVal 547
Db |||||
QY 1481 CGCGTTGTTCTCCAAACCGC---CCGTTCTTTGTG 1511
Db |||||

```

Search completed: August 5, 2004, 20:33:36  
Job time : 4255 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 5, 2004, 17:57:31 ; Search time 449 Seconds  
(without alignments)  
5175.422 Million cell updates/sec

Title: US-09-943-857-4

Perfect score: 2864

Sequence: 1 SMNSRGPAGRLGSLVPTAKLA.....DNFRAGYDALMTNPSSFFV 547

Scoring table:

BLOSUM62	Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh  
-Q=/cgn2.1/USPTO.spool/US09943857/runat.29072004.090859.29721/app.query.fasta\_1.711  
-DB=N\_Geneseq\_29Jan04\_QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -NORM=ext -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-USER=US09943857 @CGN 1.1.352 @runat.29072004.090859.29721 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

No.	Score	Query Match	Length	ID	Description
1:	Geneseq1980s:				
1:	Geneseq1980s:				
2:	Geneseq1990s:				
3:	Geneseq2000s:				
4:	Geneseq2001as:				
5:	Geneseq2001bs:				
6:	Geneseq2002s:				
7:	Geneseq2003as:				
8:	Geneseq2003bs:				
9:	Geneseq2003cs:				
10:	Geneseq2004s:				

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2864	100.0	1641	7 ABX95906	Abx95906 Candida r
2	2588	90.4	1641	7 ABX95909	Abx95909 Candida r
3	2513	87.7	1647	2 AAX33112	Aax33112 Synthetic
4	2512	87.7	1950	4 AAH43625	Aah43625 Synthetic
5	2495	87.1	1641	7 ABX95908	Abx95908 Candida r
6	2487	86.8	1649	2 AAX33111	Aax33111 Candida r
7	2434	85.0	1641	7 ABX95907	Abx95907 Candida r
8	2412	84.2	1650	2 AAT10422	Aat10422 Candida c

9	2396	83.7	1641	7 ABX95905	Abx95905 Candida r
10	1107	38.7	1828	2 AAQ54020	Aaq54020 Lipase co
11	1097	38.3	1692	2 AAQ10313	Aaq10313 Sequence
12	1088	38.0	1635	2 AAT10421	Aat10421 Geotrichu
13	1078	37.6	1674	2 AAQ05605	Aaq05605 Gene enco
14	938.5	32.8	2045	2 AAQ98578	Aaq98578 Aspergill
15	938.5	32.8	2045	2 AAT11299	Aat11299 Aspergill
16	632	22.1	8528	2 AAQ46249	Aaq46249 Construct
17	481	16.8	2089	6 ABI99790	Abi99790 Mouse isc
18	467	16.3	2544	7 ABZ81313	Abz81313 Human dru
19	462	16.1	1698	2 AAX86613	Aax86613 cDNA enco
20	462	16.1	2478	2 AAX86614	Aax86614 cDNA enco
21	452.5	15.8	1738	2 AAV21460	Aav21460 Rhodospo
22	452	15.8	1725	6 AAS17549	Aas17549 Synthetic
23	452	15.8	1752	3 ABA97180	Aba97180 Ache codi
24	452	15.8	1845	6 AAS17493	Aas17493 Human cdn
25	452	15.8	2156	7 ABZ23128	Abz23128 Nucleotid
26	452	15.8	2256	2 AAQ99002	Aaq99002 Human ace
27	452	15.8	2400	2 AAQ42496	Aaq42496 Full-leng
28	452	15.8	5767	6 AAS17547	Aas17547 Plasmid p
29	452	15.8	14446	6 AAS17548	Aas17548 Plasmid p
30	451	15.7	1725	8 ACC84169	Acc84169 Human but
31	451	15.7	1809	8 ACC84170	Acc84170 Human but
32	451	15.7	2381	6 ABN95773	Abn95773 Gene #227
33	451	15.7	2416	3 AAZ49471	Aaz49471 Human but
34	451	15.7	2416	3 AAZ49470	Aaz49470 Human wil
35	451	15.7	2416	6 AAL49277	Aal49277 Human but
36	451	15.7	2444	6 ABK32867	Abk32867 DNA encod
37	449.5	15.7	1716	4 AAH26618	Aah26618 Rhodospo
38	449.5	15.7	1967	6 AAL49276	Aal49276 Human but
39	444.5	15.5	2445	1 AAN60111	Aan60111 Sequence
40	444.5	15.5	2445	2 AAQ48393	Aaq48393 Full-leng
41	438	15.3	2424	8 ACH03894	Ach03894 Human cdn
42	438	15.3	2520	9 ADD12610	Add12610 Human ENZ
43	437.5	15.3	1470	2 AAT76597	Aat76597 Modified
44	437.5	15.3	1470	2 AAX80854	Aax80854 Para-Nitr
45	436.5	15.2	1470	2 AAX80851	Aax80851 Para-Nitr

#### ALIGNMENTS

RESULT 1

ABX95906

ID ABX95906 standard; DNA; 1641 BP.

XX

AC ABX95906;

XX

DT 15-JUL-2003 (first entry)

XX

DE Candida rugosa lipase 3 DNA.

XX

KW Lipase 3; gene; ds; mutant.

XX

OS Candida rugosa.

XX

FH Key Location/Qualifiers

FT CDS 1..1641

FT /tag= a

FT /product= "Lipase 3"

FT /partial

FT /note= "No start or stop codon shown"

XX

PN EP1288294-A2.

XX

PD 05-MAR-2003.

XX

PF 26-APR-2002; 2002EP-00009616.

XX

PR 31-AUG-2001; 2001US-00943857.

XX

PA (SINI-) ACAD SINICA.

XX

PI Tang S, Lee G, Shaw J;

XX WPI; 2003-395476/38.  
 DR P-PSDB; ABU09071.  
 XX  
 PT Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for  
 the preparation of Candida rugosa lipase for biocatalytic applications.  
 XX  
 PS Claim 25; Page 5-7; 33pp; English.  
 XX  
 CC The invention relates to an isolated mutant nucleic acid encoding a  
 CC Candida rugosa lipase polypeptide. The DNA has a sequence having at least  
 CC 80% identity to a wild-type DNA encoding Candida rugosa lipase. The  
 CC invention also relates to a microorganism comprising the DNA, where the  
 CC microorganism is a bacterium or yeast, preparing a mutant DNA encoding a  
 CC C.rugosa lipase and a chimeric C.rugosa lipase comprising a substrate  
 CC interacting domain of a first C.rugosa lipase and a non-substrate  
 CC interacting domain of a second C.rugosa lipase. The method is useful for  
 CC preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is  
 CC useful in the large scale manufacture of Candida rugosa lipase which is  
 CC useful for biocatalytic applications. This sequence represents DNA  
 CC encoding Candida rugosa lipase 3  
 XX  
 SQ Sequence 1641 BP; 315 A; 525 C; 459 G; 342 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 3, 22e-263 Length: 1641  
 Score: 2864.00 Matches: 547  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 7 Gaps: 0

US-09-943-857-4 (1-547) x ABX95906 (1-1641)

QY 1 SerMetAsnSerArgGlyProAlaGlyArgLeuGlySerValProThrAlaLysLeuAla 20  
 DB 1 TCGATGAATTCACGTGGCCAGCCGCGCTCCGATCGGTACCCACCGCCCAAGCTCGCC 60  
 QY 21 AsnGlyAspThrIleThrGlyLeuAsnAlaIleAsnGluAlaPheLeuGlyIlePro 40  
 DB 61 AACGGCGACACATCACCGGTCTCAACGCCATCATCAAGAGCGGTTCCTCGCATTTCC 120  
 QY 41 PheAlaGluProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySer 60  
 DB 121 TTTCGGACCCCGCGGTGGGCAACCTCCGCTTCAAGACCTGTGCGGTACTCTGGCTCG 180  
 QY 61 LeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGly 80  
 DB 181 CTCACGGCCAGAAAGTTCACTTCTTACGGCCGCTTTCATGCAGCAGAACCCCGAGGCG 240  
 QY 81 ThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPhe 100  
 DB 241 ACGTTTGAAGAGAACCTTGGCAAGACGGCACCTCGACTTGGTGATGAGTCCAGAGGTTC 300  
 QY 101 GlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgProProGly 120  
 DB 301 CAGGCGGTGCTTCCCGAGTGGAGTGCCTCACCATCAACGTGGTGGCGCGCGGCG 360  
 QY 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIle 140  
 DB 361 ACCAAGCGGGGCGCCAACCTCCCGGTCTGATCTCTGATCTTTGGCGGTGGTGTGATC 420  
 QY 141 GlySerProThrIlePheProProAlaGlnMetValThrLysSerValLeuMetGlyLys 160  
 DB 421 GGCAGCGCCACCATCTTCCCTCCCGCCAGATGGTCACCAAGAGTGTCTCATGGGCAAG 480  
 QY 161 HisIleIleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAsp 180  
 DB 481 CACATCATCCAGTGGCCGCTCACTACCGGTGTTGCCTCGTGGGGGTCTTGGCTGTGAT 540  
 QY 181 AspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGln 200  
 DB 541 GACATCAAGGCCGAGGCGAGCGGAAAGCCCGGCTTGAAGACACAGCGTTTGGGCATGCAG 600

QY 201 TrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePheGly 220  
 DB 601 TGGGTGGCAGACACATTGGCGGTTCGGCGGCGACCCGACGAAGGTGACTATCTTTGGC 660  
 QY 221 GluSerAlaGlySerMetSerValLeuCysHisIleuIleTrpAsnAspGlyAspAsnThr 240  
 DB 661 GAGTCTCGGGCAGCATGTCGTGTTGTGCACCTCATCTGGAACGACGGCGCAACACG 720  
 QY 241 TyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSer 260  
 DB 721 TACAAGGGCAGCCGTTGTTCCGGCGGGGATCATGAGTCTGGAGCATGGTGGCGTCT 780  
 QY 261 AspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGly 280  
 DB 781 GACCGGTGGACGCGACGATCGGCAACGAGATCTACGACCTCTTTGTCTCGAGTGTGGC 840  
 QY 281 CysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeu 300  
 DB 841 TGTGGCAGCGCCAGCGCAAGCTCGCGTGTTCGGCAGTGGCTTAGCGACACCTTGTCTC 900  
 QY 301 AspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeu 320  
 DB 901 GATGCCACCAACACATCTCTGGGTCTTGGGTACTCTCTGTTGGGTGTCTTATCTC 960  
 QY 321 ProArgProAspGlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLys 340  
 DB 961 CCGCGGCCGCGCGCAAGACATCACCGATGACATGTACAAGTTGGTGGCGACGCGCAAG 1020  
 QY 341 TyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyThrIlePheGlyLeu 360  
 DB 1021 TATGCAAGCGTTCGCGTGTATCTTGGCAGCAGACGAGGCGGACCATCTTTGGGCTC 1080  
 QY 361 SerSerLeuAsnValThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIle 380  
 DB 1081 TCTTTTGAAGTGGACACGAAAGTCTCAGCGCGGTGTACTTCAAGCAGTCTTTCTATC 1140  
 QY 381 HisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGln 400  
 DB 1141 CAGCGCAGCGCGGAGATCGACACTTGATGGCGGGGTACCCCGCAGACATCACCCAG 1200  
 QY 401 GlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSer 420  
 DB 1201 GGTTCCTCGGTTCGACACGCGGTCTCTCAACGCGCTTCAACCGCAGTTCAAGAGAATCTCT 1260  
 QY 421 AlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGln 440  
 DB 1261 GCGGTGCTCGGGACCTTGCATTTCACGCGCGCGCGCTACTTCTCAACCATTCACAG 1320  
 QY 441 GlyGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThr 460  
 DB 1321 GCGCGCACCAAGTACTCGTTCCTCTTAAGCAGCTCTCTGGGTTCGCAATCATGGGCACC 1380  
 QY 461 PheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyr 480  
 DB 1381 TTTCATGCGCAACGACATTTGTGGCAGGACTACTTGTGGGAAGCGGCGAGCTCATCTAC 1440  
 QY 481 AsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuLeuVal 500  
 DB 1441 AACACAGCGTTTATCGGTTCGCCACCGACTTGGACCCCAACACCGCGGGTGTGGTG 1500  
 QY 501 AsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAla 520  
 DB 1501 AACTGGCCCAAGTACACACGAGCTCTAGTCTGGCAACAACTTGTATGATGATCAACGCC 1560  
 QY 521 LeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAspAlaLeuMetThr 540  
 DB 1561 TTGGGCTTGTACACCGGCAAGGACAACTTCCGACCGCTGGCTACGACGCGTGTGATGACC 1620  
 QY 541 AsnProSerSerPhePheVal 547  
 DB 1621 AACCGTCTCTTCTTTCTTGTG 1641



## RESULT 2

ABX95909  
ID ABX95909 standard; DNA; 1641 BP.  
XX AC ABX95909;  
XX

XX 15-JUL-2003 (first entry)  
XX DE Candida rugosa lipase 8 DNA.  
XX

XX Lipase 8; gene; ds; mutant.  
XX OS Candida rugosa.  
XX

Key Location/Qualifiers  
FH 1. .1641  
FT CDS /tag= a  
FT /product= "Lipase 8"  
FT /partial  
FT /note= "No start or stop codon shown"  
XX

XX EP1288294-A2.  
XX

XX 05-MAR-2003.  
XX

XX 26-APR-2002; 2002EP-00009616.  
XX

XX 31-AUG-2001; 2001US-00943857.  
XX

XX (SINI-) ACAD SINICA.  
XX

XX Tang S, Lee G, Shaw J;  
XX

XX WPI; 2003-395476/38.  
XX

XX P-PSDB; ABU09074.  
XX

XX Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for  
XX the preparation of Candida rugosa lipase for biocatalytic applications.  
XX

XX Claim 25; Page 10-11; 33pp; English.  
XX

XX The invention relates to an isolated mutant nucleic acid encoding a  
XX Candida rugosa lipase polypeptide. The DNA has a sequence having at least  
XX 80% identity to a wild-type DNA encoding Candida rugosa lipase. The  
XX invention also relates to a microorganism comprising the DNA, where the  
XX microorganism is a bacterium or yeast, preparing a mutant DNA encoding a  
XX C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate  
XX interacting domain of a first C. rugosa lipase and a non-substrate  
XX interacting domain of a second C. rugosa lipase. The method is useful for  
XX preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is  
XX useful in the large scale manufacture of Candida rugosa lipase which is  
XX useful for biocatalytic applications. This sequence represents DNA  
XX encoding Candida rugosa lipase 8  
XX

SQ Sequence 1641 BP; 298 A; 537 C; 472 G; 334 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 6,89e-237 Length: 1641  
Score: 2588.00 Matches: 485  
Percent Similarity: 94.15% Conservative: 30  
Best Local Similarity: 88.67% Mismatches: 32  
Query Match: 90.36% Indels: 0  
DB: 7 Gaps: 0

US-09-943-857-4 (1-547) x ABX95909 (1-1641)

QY 1 SerMetAsnSerArgGlyProAlaGlyArgLeuGlySerValProThrAlaLysLeuAla 20

DB 1 TCGATGAATTCAGTGGCGCAGCGCGCGTCTCGGATCGGTACCCACTGCCAGCTCGCC 60

QY 21 AsnGlyAspThrIleThrGlyLeuAnAlaIleAsnGluAlaPheLeuGlyIlePro 40

DB 61 AACGGCGACCATCATCCGGTCTCAACGGCATCATCAACGAGCGGTCTCTCGGCATTTCCC 120

QY 41 PheAlaGluProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySer 60  
DB 121 TTTGCCGAGCGCGCGTGGGCAACCTCCGCTTCAAGGACCCCGTACTCTCGGCTCG 180  
QY 61 LeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGly 80  
DB 181 CTCGATGCCAGAGATTCACTTCTTACGGCCCGTCTTGATGTCAGCAGAACCCGAGGGC 240  
QY 81 ThrPheGluGlnAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPhe 100  
DB 241 ACCTACGAGGAGAACCTCCCAAGGAGCGCTCGACTTGGTGTGTCAGTCCAGGTGTTT 300  
QY 101 GlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgProGly 120  
DB 301 GAGCGGTGTCTCGTCTAGCGAGACTGTCTCACCATCAACGCTGGTGGCGCCCGGGC 360  
QY 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIle 140  
DB 361 ACCAAGCGGGTGCACACCTCCCGGTGATGCTCTGGATCTTTGGCGGGGGTTGAGGTG 420  
QY 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160  
DB 421 GGTGGCACCAACCTTCCCTCCCGCCAGATGATCACCAGAGCATTTGCCATGGCAAG 480  
QY 161 HistIleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAsp 180  
DB 481 CCCATCATCCAGTGGAGCTCAACTACCGCTGTGCTGGGGGTCTTTGGTGGCGCAC 540  
QY 181 AspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGln 200  
DB 541 GAGATCAAGGCGGAGGGCAGTGCACACCGCGGTTTGAAGACCAGCGCATGGGCATGCG 600  
QY 201 TrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePheGly 220  
DB 601 TGGGTGGCGGACAAACATTTGGCGGCTTTGGCGGCGACCGACCAAGGTGACCATTTTGGC 660  
QY 221 GluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsnThr 240  
DB 661 GAGTCTGGCGGCGAGCATGCGTCAATGTCACATTTCTGGAAACACGCGCGCAACACG 720  
QY 241 TyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSer 260  
DB 721 TACAAGGCGCAAGCGCTCTTCGCGCGGCGCATCATGCACTCTGGGCGCATGTGACGTGCG 780  
QY 261 AspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGly 280  
DB 781 GACGCGTGGACGCGCTCTACGCGCAACGAGATCTTTGACCTCTTGGCGTCGCGACGCGGC 840  
QY 281 CysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeu 300  
DB 841 TCGCGGCGCGCGACGCGACAGCTTGGCTGTCTGGCGGGTGTGTCTAGCGACACGCTTGAG 900  
QY 301 AspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeu 320  
DB 901 GACGCCACCAACAACACCTCGGTTCTTGGCGTACTCTCTCGTTGCGGTGCTTTATCTC 960  
QY 321 ProArgProAspGlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLys 340  
DB 961 CGCGCGCGCGCGGTGAACATCAACGCGACGACATGTTTGGCTTGTCTCGCGCGGCGCAAG 1020  
QY 341 TyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyThrIlePheGlyLeu 360  
DB 1021 TATGAAGCGTTCTGTGTATCATCGCGCACCAAGACGAGCGGCGCACCTTTCTTGGCAC 1080  
QY 361 SerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIle 380  
DB 1081 TCTTTCTTTGAACGTGACCAAGATCGCGGAGCGCGCGAGTACTTTCAGCAGTCTTTTGTG 1140  
QY 381 HisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGln 400  
DB 1141 CACGCGCAGCGCGAGCTCGACACGTTGATGAGCGGGTACCCCCAGGACATCACCCAG 1200

Fri Aug 6 10:53:11 2004

us-09-943-857-4.rng

QY 401 GlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSer 420  
 Db 1201 GGTTCCTCGTTCGACACGGGTGTCTCAACGCCCTCACCCCGCAGTTCACAGAGATCTCT 1260  
 QY 421 AlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGln 440  
 Db 1261 CGGTGTCTCGCGACCTTGTCTTCATCCAGCCGCTGCTACTTCTCTCAACCACTACAC 1320  
 QY 441 GlyGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThr 460  
 Db 1321 GCGGACCAAGTACTCATCTCTTAAGCAGCTCTCTGGCTTGGCGGTCTCGAAGC 1380  
 QY 461 PheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyr 480  
 Db 1381 TTCCACTCCACGACATGTCTTCAGGACTACTTGTGGCAGCGGTCTCTCATCTAC 1440  
 QY 481 AsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuLeuVal 500  
 Db 1441 AACAAAGCGTTCATGTGCTTGGCCAGGACTTGGACCCCAACACCGCGGGTGTGTGTG 1500  
 QY 501 AsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAla 520  
 Db 1501 AAGTGGCCCGAGTACACAGCAGCTCTCAGTCTGGCAACAATGATGATGATCAACGCC 1560  
 QY 521 LeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAspAlaLeuMetThr 540  
 Db 1561 TTGGCTTTGTACACCGCAAGACAACCTCCCGCACCGCGGTACGACGCGTGTCTCC 1620  
 QY 541 AsnProSerSerPhePheVal 547  
 Db 1621 AACCCGCGCTCTTCTTTGTG 1641

RESULT 3  
 ID AAX33112 standard; DNA; 1647 BP.

XX AAX33112;  
 XX 23-JUN-1999 (first entry)

Synthetic lipase 1 gene.

Candida rugosa; lipase 1; LIP1; industrial bioconversion; ss.

Synthetic.  
 Candida rugosa.

WO9914338-A1.

25-MAR-1999.

16-SEP-1997; 97WO-NL000524.

16-SEP-1997; 97WO-NL000524.

(UNIL ) UNILEVER NV.

Brocca S, Schmidt-Dannert C, Lotti M, Alberghina L, Schmid R;  
 WPI; 1999-229539/19.

Synthesis and functional overexpression of a Candida rugosa lipase gene  
 coding for a major industrial lipase.

Disclosure; Page 23-28; 44pp; English.

The present sequence represents a synthetic lipase 1 gene derived from  
 the native Candida rugosa lipase 1 gene. Lipases produced by Candida  
 rugosa are extensively used in industrial bioconversions, and the pure  
 lipase 1 can be used in a process requiring high specificity toward acyl  
 chains shorter than 14C. Lipase 1, free of 2-5, can be obtained without  
 using expensive and expensive working up procedures. Pure lipase 1  
 exhibits higher activity toward capriate than toward palmitate

XX SQ Sequence 1647 BP; 377 A; 400 C; 364 G; 506 T; 0 U; 0 Other;  
 Alignment Scores: 9.9e-230 Length: 1647  
 Pred. No.: 2513.00 Matches: 470  
 Score: 93.48% Conservative: 32  
 Percent Similarity: 87.52% Mismatches: 35  
 Best Local Similarity: 87.74% Indels: 0  
 Query Match: 2 Gaps: 0  
 DB: 1647  
 US-09-943-857-4 (1-547) x AAX33112 (1-1647)  
 QY 11 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30  
 Db 37 GTTCTGCTGCTGCCCCAACCGGCACCTTGGCTAAACGGTGACACCATCCCGGTTTGAAGCC 96  
 QY 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArg 50  
 Db 97 ATCATCAACGAAGCCCTTCTTGGGTATTCCATTTGCCGAACCCAGCTTGGTAACCTTGAGA 156  
 QY 51 PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly 70  
 Db 157 TTCAGGACCCAGTTCCTATCTCCGTTCTTGGATGGTCAAAAGTTCCATCTCTTACGGT 216  
 QY 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90  
 Db 217 CCATCTGTATGCAACAAACCCAGAGGTACTACGAAGAAACTTCCCAAGACGCT 276  
 QY 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110  
 Db 277 TTAGATCTGGTATTGCAATCCAAAGTTTTCGAAGCTGTTCCTCATCTCTTGAAGACTGT 336  
 QY 111 LeuThrIleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeuProValMet 130  
 Db 337 TTGACCATTAATGTGTAGACCACCGCGACAAAGGTGGTGCCAACTTGCACGTATG 396  
 QY 131 LeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePheProAlaGln 150  
 Db 397 TTGTGGATCTTTGGTGGTGGTGTTCGAAGTTGGTGTACTAGTACTCTCCCTCCAGCCCAA 456  
 QY 151 MetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArg 170  
 Db 457 ATGATTACCAAGTCTATTGCTATGGTAAGCAATCAATCCAGCTTCTGTCACTACAGA 516  
 QY 171 ValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAla 190  
 Db 517 GTCCTGAGCTGGGGTTTCTTGGCTGGTACGAATCAAGCGCGAAGTTCTGCCAACGCC 576  
 QY 191 GlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly 210  
 Db 577 GGTTTGAAGGACCAAGATTGGGTATGCAATGGGTGGCTGACAACTTGTCTTGTGTGT 636  
 QY 211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230  
 Db 637 GGTGATCCAACTAAAGGTACTATCTTTGGTGAATCTGCTGGTCTATGTCCTGCTCACTGT 696  
 QY 231 HisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250  
 Db 697 CACATTTTGGGAACGCGGTGACACACTTACAAAGGGTAAGCAATGTTTCAGAGCTGT 756  
 QY 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270  
 Db 757 ATCATGCAATCTGGTGTATGGTTCATCTGACGCGGTGACGGTATCTACGGTAACGAA 816  
 QY 271 IleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290  
 Db 817 ATTTTTGAATCTTGTGGCTTCCACGCTGGTGTGGTGTCTGCTCTGCAAGTTGGCTGT 876  
 QY 291 LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310  
 Db 877 TTGAGAGGTGTTTCTTCTGACACTTGGGAAGACGCCCAACCAACACCCCTGGTCTTGT 936



QY	133	IlePheGlyClyGlyPheGluIleGlySerProThrIlePheProAlaGlnMetVal	152
DB	532	ATCTTTGGTGGTGGTTTTGAAGTTGGTGTAAGTAGTCCTTCCCTCACGCCAATGATT	591
QY	153	ThrlvsSerValLeuMetGlyLysHisIlelleHlsValAlaValAsnTyrArgValAla	172
DB	592	ACCAAGTCTATTGCTATGGGTAAAGCCAATCATCCAGTTTCTGTCAACTACAGAGTCTCG	651
QY	173	SerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGlnGlySerGlyAsnAlaGlyLeu	192
DB	652	AGCTGGGGTTCTTGGCTGGTGACAAATCAAGGCCGAAGGTTCTGCCAACCGCGTTTG	711
QY	193	LysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAsp	212
DB	712	AAGGACCAAGAATTGGTATGCATGGTGGCTGCACAACATTCGCTTTGGTGGTGTAT	771
QY	213	ProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeu	232
DB	772	CCAAATAAGGTTACTATCTTTGGTGAATCTGCTGGTTCTATGTCCGTATGTGTCCAT	831
QY	233	IleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMet	252
DB	832	TTGTGGAACACGCGTGACAACTTACAGGGTAAGCCATTGTTTCAGAGCTGGTATCAG	891
QY	253	GlnSerGlyAlaMetValProSerAspProValaspGlyThrTyrGlyAsnGluIleTyr	272
DB	892	CAATCTGGTGTATGGTTCCATCTGCAGCCGTCGACGGTATCTACGGTAACGAAATTTT	951
QY	273	AspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArg	292
DB	952	GACTTGTGGCTTCCAAACGCTGGTGTGGTTTGCCTCTCACAGTGGGCTTTGTTGAGA	1011
QY	293	SerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyr	312
DB	1012	GGTGTTTCTCTGCACCTTTGGAGAAGCCACCACCAACACCCCTGGTTCTTGGCTTAC	1071
QY	313	SerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAnilleThrAspAspMet	332
DB	1072	TCTCTCTTAAGATTGTCTTACTTCCCAAGACACAGCGGTGTAAACATPCACCGACGACATG	1131
QY	333	TyrllysLeuValArgAspGlyLysTyrAlaSerValProValIlelleGlyAspGlnAsn	352
DB	1132	TACGCTTTGGTTAGAGAAGGTAAAGTAGCCCAACATCCCTGTATCATCGTGCACCAAAC	1191
QY	353	AspGluGlyThrIlePheGlyLysSerSerLeuAsnValThrThrAsnAlaGlnAlaArg	372
DB	1192	GACGAAGGTACCTCTTTGGTACTTCTTCTTTGAACCTTACCACTGATGCCAACGCCAGA	1251
QY	373	AlaTyrrPheLysGlnSerPheIleHlsAlaSerAspAlaGluIleAspThrLeuMetAla	392
DB	1252	GAATATTTCAAGCAATCTTTTGTCCACGCTAGCAGCGCTGAAATTCGACACTTTGATGACT	1311
QY	393	AlaTyrrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeu	412
DB	1312	GCTTACCCAGGTGACATCACTCAAGTTCTCCATTGGACATCGGAATCTAAACGCCTTG	1371
QY	413	ThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHlsAlaArg	432
DB	1372	ACCCACACAATTCAGAGAAATCTCTGCTTTTGGGTGACTTGGGGTTTACTTTGGCTCGT	1431
QY	433	ArgTyrrPheLeuAsnHisPheGlnGlyGlyThrLysTyrSerPheLeuSerLysGlnLeu	452
DB	1432	AGATACTTCTTGAACCACTACACCGGTGGTACCAAGTACTCTTCTTGTCTTAAGCAATTG	1491
QY	453	SerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrIleu	472
DB	1492	TCTGGTTTGGCAGTTTGGGTACTTCTCCACTCCAACGATATCGTCTTCCAAAGACTACTTG	1551
QY	473	LeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAsp	492
DB	1552	TTGGGTTCTGGTTCCTTGATCTACAAACGCTTTTCATTTGCTTTGGCCACTACTTGAC	1611

Qy	493	ProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGlnSerGly	512
Db	1612	CCAAACACCCCGGTTTGGTTAAAGTGGCCAGAAATACACCTCTCTTCTCAATCTGGT	1671
Qy	513	AsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThr	532
Db	1672	AACAACCTTGATGATGATCAACGCTTTGGGTTTGACACCGGTAAAGACAACATTCAGAAC	1731
Qy	533	AlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal	547
Db	1732	GCCGGTTACACGCTTGTCTCCACCCACCATCTTCTTTGTT	1776
RESULT 5			
ID	ABX95908	standard; DNA; 1641 BP.	
AC	ABX95908;		
XX			
DT	15-JUL-2003	(first entry)	
XX			
DE	Candida rugosa lipase 5 DNA.		
XX			
KW	Lipase 5; gene; ds; mutant.		
XX			
OS	Candida rugosa.		
XX			
XX			
FF	Key	Location/Qualifiers	
FT	CDS	1..1641	
FT		/*tag= a	
FT		/product= "Lipase 5"	
FT		/partial	
FT		/note= "No start or stop codon shown"	
XX			
PN	EP1289294-A2.		
XX			
PD	05-MAR-2003.		
XX			
PF	26-APR-2002; 2002EP-00009616.		
XX			
PR	31-AUG-2001; 2001US-00943857.		
XX			
PA	(SINI-) ACAD SINICA.		
XX			
PI	Tang S, Lee G, Shaw J;		
XX			
DR	WPI; 2003-395476/38.		
DR	P-P5DB; ABU09073.		
XX			
PT	Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for the preparation of Candida rugosa lipase for biocatalytic applications.		
XX			
PS	Claim 25; Page 8-9; 33pp; English.		
XX			
CC	The invention relates to an isolated mutant nucleic acid encoding a Candida rugosa lipase polypeptide. The DNA has a sequence having at least 80% identity to a wild-type DNA encoding Candida rugosa lipase. The invention also relates to a microorganism comprising the DNA, where the microorganism is a bacterium or yeast, preparing a mutant DNA encoding a C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate interacting domain of a first C. rugosa lipase and a non-substrate interacting domain of a second C. rugosa lipase. The method is useful for preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is useful in the large scale manufacture of Candida rugosa lipase which is useful for biocatalytic applications. This sequence represents DNA encoding Candida rugosa lipase 5		
XX			
SQ	Sequence 1641 BP; 302 A; 543 C; 444 G; 352 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	5,14e-228	Length:	1641
Score:	2495.00	Matches:	472
Percent Similarity:	91.96%	Conservative:	31
Best Local Similarity:	86.29%	Mismatches:	44

Query Match:	87.12%	Indels:	0
DB:	7	Gaps:	0
US-09-943-857-4 (1-547) x ABX95908 (1-1641)			
QY	1	SerMetAsnSerArgGlyProAlaGlyArgLeuGlySerValProThrAlaLysLeuAla	20
DB	1	TCATGAATTCAGGTGGCCAGGCGCGGTCTCGGATCGGTACCCCACTGCCAGCGTCGCC	60
QY	21	AsnGlyAspThrIleThrGlyLeuAsnAlaIleleasngluAlaPheLeuGlyIlePro	40
DB	61	ACGGCGACACCATCACCGGTCTCAACGCCATCATCAACGAGGCGTTCCTCGGCATTCCC	120
QY	41	PheAlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySer	60
DB	121	TTTGCCGAGCGCGGTGGCAACCTCCGCTTCAAGGACCCCTGTGCGGTACCGTGGGTCT	180
QY	61	LeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGly	80
DB	181	CTCAACGGTCAATCTCTTCAACCGGTCAGGTCCTGCTTTCATGATGAGCAGAACCCCGAGGC	240
QY	81	ThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPhe	100
DB	241	ACCTACGAGGAGACCTCCCAAGGTGGCGCTTGACTTGGTATGAGTCCAAAGTGTTTC	300
QY	101	GlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgProProGly	120
DB	301	CAGGCTGTCTCTCCCAACAGCAGGAGACTGCCTCACCATCAACGTGGTGGCGCGCGGC	360
QY	121	ThrLysAlaGlyAlaAsnLeuProValMetLeuThrIlePheGlyGlyPheGluIle	140
DB	361	ACCAAGCGCGCGGCAACCTCCCGGTCATGCTCTGGATCTTTGGCGGTGGGTTTGATC	420
QY	141	GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys	160
DB	421	GGCAGCCCCACCATCTTCCCTCCGCTCAGATGCTTCCNAGAGTGTCTCATGGCGAG	480
QY	161	HisIleIleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAsp	180
DB	481	CCCATCATCCAGTGGCGCTCAACTACCGCTTGGCGTCTTTGGTTCTTGGCGGTCGC	540
QY	181	AspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGln	200
DB	541	GACATCAAGCGCGAGGCGAGCTCCAAATGGCGGCTCAAGSACCAGCGCTTGGCGATGAG	600
QY	201	TrpValAlaAspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePheGly	220
DB	601	TGGTGGCAGACNACATTCGCGGTTTCGCGGCGACCCGAGCAAGTGACCATCTTTGGC	660
QY	221	GluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsnThr	240
DB	661	GAGTCTGCGCGGAGCATGTCGCTGTGTGTCACCTTCTTGGATGGCGGCGACACACG	720
QY	241	TyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSer	260
DB	721	TACAAGGCAAGCGGTGTTCGCGCGGCGCATCATGCTCTGGAGCCATGTCGCGCTCT	780
QY	261	AspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGly	280
DB	781	GACCGGTGGACGGACCTATGAGGCCAAATCTATGACAGTGTGGTGGCTTCTACGGGC	840
QY	281	CysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeu	300
DB	841	TGCAGCAGTGCAGCAACAAGCTTGGCTTGGCTGCTTCTTACTCAGGCATTTGCTC	900
QY	301	AspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeu	320
DB	901	GATGCCACCAACACACCCCTGGTCTTGTGTATACACCTCGTGGGTGTTCTTATCTC	960
QY	321	ProArgProAspGlyLysAsnIleThrAspMetTyrLysLeuValArgAspGlyLys	340
DB	961	CGCGGCGCGAGCGGCCCAACATCACCGATGACATGTACAAGTTGTGTCGCGAGCAAG	1020
QY	341	TyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyThrIlePheGlyLeu	360
DB	1021	TATGCAAGCGTTCCTCGGTGATCATTTGGGACCAACGACGAGGCTCTCTTTGTGATCTC	1080
QY	361	SerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIle	380
DB	1081	TCCTCTTTGAACACCAACCGAGGCGGACCGCGAGCATACCTCAGAAAGTCTTTTCATC	1140
QY	381	HisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGln	400
DB	1141	CACGCCACCGACCCGATATCACCGCATTAAGCGGCGGTACCCACGAGCATGTCACCCAG	1200
QY	401	GlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSer	420
DB	1201	GGTTCCTCGGTTCGACACGGGCATCTCAAGCCCTTACACCCAGCTCAAGCGGATCAAT	1260
QY	421	AlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGln	440
DB	1261	GCTGTGCTTGGCGACCTTACCTTACCTCTCTAAGCAGCTTCTGGGTGGCCATTCTCGGTACG	1320
QY	441	GlyGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThr	460
DB	1321	GGTGTGCTTGGCGACCTTACCTTACCTCTCTAAGCAGCTTCTGGGTGGCCATTCTCGGTACG	1380
QY	461	PheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyr	480
DB	1381	TTCCACGCGAAGCAGCATTTGTGGCAGCACTTTTGTGGCAGCGCGAGCGTCATCTAC	1440
QY	481	AsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuLeuVal	500
DB	1441	AACAACGCGTTCATCGCGTTTGCACCGCACTTGGACCCCAACACCGCGGCTTGTCTGTG	1500
QY	501	AsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAla	520
DB	1501	CAGTGGCCCCAGTACACAGCAGCTCTCAGCGGGGAGCAACTTGTATGAGATCAGTGCC	1560
QY	521	LeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAspAlaLeuMetThr	540
DB	1561	TTGGGCTTGTACACCGGCAAGCAACACTTCGCGCGCGGCTACCAACGCTTGTGTGCC	1620
QY	541	AsnProSerSerPhePheVal	547
DB	1621	GACCGCTCTCACTTTTTCGTG	1641
RESULT 6			
ID	AA33111	standard; DNA; 1649 BP.	
AC	AA33111;		
DT	23-JUN-1999	(first entry)	
DE	Candida rugosa lipase 1 gene.		
DE	Candida rugosa; lipase 1; LIP1; industrial bioconversion; ss.		
OS	Candida rugosa.		
PN	WO9914338-A1.		
XX	25-MAR-1999.		
XX	16-SEP-1997; 97WO-NL000524.		
XX	16-SEP-1997; 97WO-NL000524.		
XX	(UNIL ) UNILEVER NV.		
PI	Brocca S, Schmidt-Bannert C, Lotti M, Alberghina L, Schmid R;		
DR	WPI; 1999-229539/19.		
PT	Synthesis and functional overexpression of a Candida rugosa lipase gene		



```

FH Key      Location/Qualifiers
FT CDS      1..1641
FT FT       /*tag= a
FT FT       /product= "Lipase 4"
FT FT       /partial
FT FT       /note= "No start or stop codon shown"
XX
PN BP1288294-A2.
XX
XX 05-MAR-2003.
XX
XX 26-APR-2002; 2002EP-00009616.
XX
XX 31-AUG-2001; 2001US-00943857.
XX (SINI-) ACAD SINICA.
XX
XX Tang S, Lee G, Shaw J;
XX
XX WPI; 2003-395476/38.
XX P-PSDB; ABU09072.
XX
XX Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for
XX the preparation of Candida rugosa lipase for biocatalytic applications.
XX
XX Disclosure; Page 7-8; 33pp; English.
XX
XX The invention relates to an isolated mutant nucleic acid encoding a
XX Candida rugosa lipase polypeptide. The DNA has a sequence having at least
XX 80% identity to a wild-type DNA encoding Candida rugosa lipase. The
XX invention also relates to a microorganism comprising the DNA, where the
XX microorganism is a bacterium or yeast, preparing a mutant DNA encoding a
XX C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate
XX interacting domain of a first C. rugosa lipase and a non-substrate
XX interacting domain of a second C. rugosa lipase. The method is useful for
XX preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is
XX useful in the large scale manufacture of Candida rugosa lipase which is
XX useful for biocatalytic applications. This sequence represents DNA
XX encoding Candida rugosa lipase 4
XX
XX Sequence 1641 BP; 306 A; 524 C; 461 G; 350 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 3..39e-222 Length: 1641
XX Score: 2434.00 Matches: 457
XX Percent Similarity: 90.49% Conservative: 38
XX Best Local Similarity: 83.55% Mismatches: 52
XX Query Match: 84.99% Indels: 0
XX DB: 7 Gaps: 0
XX
XX US-09-943-857-4 (1-547) x ABX95907 (1-1641)
XX
QY 1 SerMetAsnSerArgGlyProAlaGlyArgLeuGlySerValProThrAlaLysLeuAla 20
DB 1 TCATGAATTACGTGGCCAGCCGCGCTCTCGATCGGTATCCCACTGCCACGCTCGCC 60
QY 21 AsnGlyAspThrIleThrGlyLeuAsnAlaIleAsnGluAlaPheLeuGlyIlePro 40
DB 61 AACGGCGACACCATCACCGGTCTCAACGCCATCATCAACAGCGGTCTCTCGTATATCCC 120
QY 41 PheAlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySer 60
DB 121 TTGTGCTCAGCGCGCGGTGGGCAACCTCCGCTTCAAGCGGCTGTGCGGTACTCGGCGTCT 180
QY 61 LeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGly 80
DB 181 CTCAATGGTCAAGATTACTTCGTATGGCCCTTCGTGATCGATGATGAACCCATTGGGC 240
QY 81 ThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPhe 100
DB 241 AACTGGGACTCTCGCTTCCCAAGGCTGCCATCACTCTTGATGAGTCCCAAGCTCTTC 300
QY 101 GlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgProGly 120

```

Db 1381 CACCACGCCACGACATTTGTGGCAGGACTTTTGTGGTGGCCACAGCAGCGCGTGTAC 1440  
 Qy 481 AsnAenAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuVal 500  
 Db 1441 AACAAACGGTTATTGGCTTTGGCCAAACGCTCGACCGCAAGCGGGTTTGTGTG 1500  
 Qy 501 AsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMetIleAsnAla 520  
 Db 1501 AACTGGGCCAAAGTACACAGCAGCTCTCAGTCAGGCAACAATTTGTTCAGATCAACGCC 1560  
 Qy 521 LeuGlyLeuTyrThrGlyLeuAspAsnPheArgThrAlaGlyTyrAspAlaLeuMetThr 540  
 Db 1561 TTGGCTGTATACCGGCAAGGACAACTTCGCGACCGTGTGTAGACGCGTTGTTTACC 1620  
 Qy 541 AsnProSerSerPheVal 547  
 Db 1621 AACCGCTCTCTCTCTTTGTG 1641

RESULT 8

AT10422  
 ID AT10422 standard; DNA; 1650 BP.

AC XX

AT10422;

DT 20-SEP-1996 (first entry)

XX Candida cylindracea lipase gene.

DE

XX Fatty acid; transgenic plant; exogenous; lipase; Rhizopus niveus;  
 KW Pseudomonas aeruginosa; Pseudomonas fluorescens; Geotrichum candidum;  
 KW Candida cylindracea; constitutive; tissue specific; promoter; lipid;  
 KW milling; biofuel; lubricant; detergent; ss.

XX Candida cylindracea.

XX Key

XX Location/Qualifiers

FT 46..1647

FT /tag= a

FT /product= "lipase enzyme"

XX PR2722798-Al.

XX 26-JAN-1996.

XX 25-JUL-1994; 94FR-00009272.

XX 25-JUL-1994; 94FR-00009272.

XX (NAPO-) INST NAT POLYTECHNIQUE TOULOUSE.

XX Alibert G, Mouloungui Z, Boudet A;

XX WPI; 1996-107680/12.

XX Prodn. of fatty acids or derivs. from transgenic oilseed plants -  
 PT engineered to express a lipase that contacts lipid(s) only when seeds are  
 PT milled.

XX Claim 9; Page 26-27; 32pp; French.

XX A novel method of producing fatty acids or their derivs. in a plant  
 CC comprises generating a transgenic plant contg. an exogenous lipase gene,  
 CC esp. selected from the Rhizopus niveus, Pseudomonas aeruginosa,  
 CC P.fluorescens, Pseudomonas sp., Geotrichum candidum or Candida  
 CC cylindracea lipase genes (AA10417-22 resp.). The lipase genes can be  
 CC placed under control of a constitutive or a tissue specific promoter. The  
 CC production of the fatty acids only occurs when the lipase and lipids  
 CC contact each other after milling of the plants. The fatty acids generated  
 CC can be used to prod. e.g. biofuels, lubricants, detergents, etc

XX Sequence 1650 BP; 299 A; 540 C; 491 G; 320 T; 0 U; 0 Other;

SQ

Alignment Scores:

Pred. No.: 4.29e-220 Length: 1650  
 Score: 2412.00 Matches: 453  
 Percent Similarity: 90.50% Conservative: 33  
 Best Local Similarity: 84.36% Mismatches: 51  
 Query Match: 84.22% Indels: 0  
 DB: 2 Gaps: 0  
 US-09-943-857-4 (1-547) x AAT10422 (1-1650)  
 Qy 11 LeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAla 30  
 Db 37 GTGGCTGTGCCCCACCGCCACCTTCCCAACCGCGACACCATCATCCGGTCTCAACGCC 96  
 Qy 31 IleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArg 50  
 Db 97 ATCATCAACGAGGCGTCTCTCGGCATTCCTTTCGAGCGCGCGTGGCAACCTCCGC 156  
 Qy 51 PheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGly 70  
 Db 157 TTCAGGACCCCGTCCGCTACTCCGGCTCGTTCGATGCCAGAGTTTCACGCTGTACCGC 216  
 Qy 71 ProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAla 90  
 Db 217 CGCTGTGCATGCAGCAGAACCCCGAGGGCCTACGAGGAGAACCTCCCAAGCAGCG 276  
 Qy 91 LeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCys 110  
 Db 277 CTCGACTTGGTGATGCAGTCCAAAGGTGTTGAGCGCGTGTCTGCCCTGAGCGAGACTGT 336  
 Qy 111 LeuThrIleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeuProValMet 130  
 Db 337 CTCACCATCAACGTCGCGCGCGCGCGCCAAAGCGGGTGCACACTCCCGGTGATG 396  
 Qy 131 LeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePheProProAlaGln 150  
 Db 397 CTCGTGATCTTTGGCGCGGGTTTGGGTGGGTGGCCACGACACCTTCCCTCCCGCCAG 456  
 Qy 151 MetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArg 170  
 Db 457 ATGATCACCAGAGCATTTGCCATGGGCAAGCCCATCATCCAGTGAAGGTCAACTACCGC 516  
 Qy 171 ValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGlyLysSerGlyAsnAla 190  
 Db 517 GTGTCGTGTGGGGGTCTTGTGCTGCGCAGAGATCAAGCGGAGCGGAGTCCCAACGCC 576  
 Qy 191 GlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGly 210  
 Db 577 GGTTTGAGGACCAAGCGCTTGGGCATGCAGTGGTGGTGGCGGACACATTCGCGGCTTTGC 636  
 Qy 211 GlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCys 230  
 Db 637 GCGGACCCGACCCAGGTGACCATCTTTGGCGAGCTGGCGGCGGAGCATGTGCGTCAATGTC 696  
 Qy 231 HisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGly 250  
 Db 697 CACATTTCTTGAACGACGCGGACACACGTAACAGGCGAAGCGGCTCTTCGCGCGGCG 756  
 Qy 251 IleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGlu 270  
 Db 757 ATCATGCAGCTGGGGGCCCATGTTGCGCTGGACGCGCTGGACGCGCATCTACGCAACGAG 816  
 Qy 271 IleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCys 290  
 Db 817 ATCTTTGACCTCTTGGCGTCAACGCGGCTCGCGCAGCGCGGCGGCGGAGCTTGGCTGC 876  
 Qy 291 LeuAspSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310  
 Db 877 TTGCGCGGTGTGCTGAGCGACAGCTTGGAGGACGCCACCAACACACCCCTGGTTCTTG 936  
 Qy 311 AlaTyrSerSerIleArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAsp 330  
 Db 937 GCGTACTCTCGTTGCGGTGTGTTGTTACTCTCCCGCGCGCGGCGGTGAACATCACCGAC 996



```

QY 331 AspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAsp 350
Db 997 GACATGTACCGCTTGTGGCGGCAAGGCAAGTATGCCAACATCCCTGTGATCATCGCGGAC 1056
QY 351 GluAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrAsnAlaGln 370
Db 1057 CAGAACACGAGGAGCCTTCTTTGGCACCCCTGCTGTGAACGTGACCGCATGCCGAG 1116
QY 371 AlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeu 390
Db 1117 GCCCGGAGTACTTCAAGACGCTGTGTTCACGCCAGCGAGCGGAGATCGACAGTTG 1176
QY 391 MetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsn 410
Db 1177 ATGACGGCGTACCCCGGCGATACACCGCGGCTGCGGCTCGACACGGGTATTCTCAAC 1236
QY 411 AlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHis 430
Db 1237 GCCCTCACCCCGGAGTTCAAGAGAATCCTGGCGGTGCTCGCGGACCTTTAGCGTT 1296
QY 431 AlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLys 450
Db 1297 GCTCGTGCCTACTTCTCAACCACTACACCGCGGCGACCAAGTACTCTATCTCTCTGAAG 1356
QY 451 GlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAsp 470
Db 1357 CAGCTCTCGGCTTGGCGGTGCTCGGAACGTTCCACTCCCAACGACATTGTCTCCAGGAC 1416
QY 471 TyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490
Db 1417 TACTTGTGGGACGCGCTCGCTCATCTCAACAACGCGTTCATTGGCGTTTGGCACGGAC 1476
QY 491 LeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db 1477 TTGGACCCCAACACCGCGGGTGTGTGTGAAGTGGCCCGAGTACACCACGACCGCTGCAG 1536
QY 511 SerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPhe 530
Db 1537 CTGGGCAACAACCTGATGATGATCAACGCTTGGGCTTGTACACCGCGCAAGGACAACTTC 1596
QY 531 ArgThrAlaGlyTyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db 1597 CGCACCGCGGCTACGACGCGTGTGTCTCCAAACCGCGCGCTGTCTTTGTG 1647

RESULT 9
ABX95905
ID ABX95905 standard; DNA; 1641 BP.
XX AC ABX95905;
XX 15-JUL-2003 (first entry)
XX DE Candida rugosa lipase 2 DNA.
XX KW Lipase 2; gene; ds; mutant.
XX OS Candida rugosa.
XX Key Location/Qualifiers
CDS 1..1641
FT /*tag= a
FT /product= "Lipase 2"
FT /partial
FT /note= "No start or stop codon shown"
XX PN EP1288294-A2.
XX PD 05-MAR-2003.
XX PF 26-APR-2002; 2002EP-00009616.
XX PR 31-AUG-2001; 2001US-00943857.
XX

```

PA (SINI-) ACAD SINICA.  
 XX Tang S, Lee G, Shaw J;  
 XX WPI; 2003-395476/38.  
 DR P-PSDB; ABU09070.  
 XX  
 PT Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for  
 PT the preparation of Candida rugosa lipase for biocatalytic applications.  
 XX  
 PS Claim 25; Page 4-5; 33pp; English.  
 XX

The invention relates to an isolated mutant nucleic acid encoding a  
 CC Candida rugosa lipase polypeptide. The DNA has a sequence having at least  
 CC 80% identity to a wild-type DNA encoding Candida rugosa lipase. The  
 CC invention also relates to a microorganism comprising the DNA, where the  
 CC microorganism is a bacterium or yeast, preparing a mutant DNA encoding a  
 CC C.rugosa lipase and a chimeric C.rugosa lipase comprising a substrate  
 CC interacting domain of a first C.rugosa lipase and a non-substrate  
 CC interacting domain of a second C.rugosa lipase. The method is useful for  
 CC preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is  
 CC useful in the large scale manufacture of Candida rugosa lipase which is  
 CC useful for biocatalytic applications. This sequence represents DNA  
 CC encoding Candida rugosa lipase 2

SQ Sequence 1641 BP; 305 A; 521 C; 472 G; 343 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 1,43e-218 Length: 1641  
 Score: 2396.00 Matches: 450  
 Percent Similarity: 88.48% Conservative: 34  
 Best Local Similarity: 82.27% Mismatches: 63  
 Query Match: 83.66% Indels: 0  
 DB: 7 Gaps: 0

US-09-943-857-4 (1-547) x ABX95905 (1-1641)

QY 1 SerMetAsnSerArgGlyProAlaGlyArgLeuGlySerValProThrAlaLysLeuAla 20  
 Db 1 TCGATGAATTCACGTGGCCCGCGCGCGCTCGGATCGGTATCCACCGCCGCTCGCC 60  
 QY 21 AsnGlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGluAlaPheLeuGlyIlePro 40  
 Db 61 AACGGCGACACCATCACCGGTCTCAACGCCATTGTCAAGAAAGTTTCTCGCATACCG 120  
 QY 41 PheAlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySer 60  
 Db 121 TTGCGGAGCGCGCGCGTGGCGACCTCCGCTTCAAGCGCGCGCTACTCGGCGTCG 180  
 QY 61 LeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGly 80  
 Db 181 CTCACGCGCGCAGGATTTACCTCTTACGGCCCGCTTGTGATGATGATGAACCTATGGGC 240  
 QY 81 ThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPhe 100  
 Db 241 TCGTTTGAGGACACACTTCCCAAGAAATGCGTTGACTTGGTGTCTCCAGTCCAAGATCTTC 300  
 QY 101 GlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgProProGly 120  
 Db 301 CAAGTGTGTCTTCCCAACACGACGAGGACTGTCTCACCATCAAGTGTATCGCGCGCGCGC 360  
 QY 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIle 140  
 Db 361 ACCAGGCGCAGTGTGCTCTCCCGGTGATGTCTTGGATCTTTGGCGGTGGGTGTGAGCTT 420  
 QY 141 GlySerProThrIlePheProProAlaGlnMetValThrLysSerValLeuMetGlyLys 160  
 Db 421 GGCGGCTCCGCGCTCTTCCAGGAGACAGATGGTGGCCAAAGCGGTGCTCATGGGTAA 480  
 QY 161 HisIleIleHisValAlaValAsnTyrArgValAlaSerTyrGlyPheLeuAlaGlyAsp 180  
 Db 481 CCGGTGATCCAGTCAGCATGAATACCGCGTGGGTCATCGGGGTCTCTTGGCGCGCGCC 540

181 AspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGln 200  
 541 GACATCCAGAACGAGGAGCGGGAACCCCGCTTGTGATGACCGAGCGGTGGCCATCGAG 600  
 201 TrpValAlaAspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePheGly 220  
 601 TGGGTGGCGGACACATTTGCTGGGTTGGCGCGACCGGACGACAGGTGACCATATACGC 660  
 221 GluSerAlaGlySerMetSerValLeuGlyCysHisLeuIleTrpAsnAspGlyAspAsnThr 240  
 661 GAGTCTGGCGGACGATGTCGACGTTTGTGACCTTGTGTGAAACGACGCGGACACACG 720  
 241 TyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSer 260  
 721 TACAACGCGCAGCGGTTTTCGCGCCGCGCATGCGAGTCTGGGTGATGGTGGCGTCT 780  
 261 AspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGly 280  
 781 GACCGCGTGGACGCGGACGTCAGCGACCGGAGTCTACAACCGAGGTGCTGGCGGTCTGCCGG 840  
 281 CysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeu 300  
 841 TGTGCGAGTGGCGGACGAGCTCGCTGCTTGGCGGCGCTTCTCAGGACACGTTGTAC 900  
 301 AspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeu 320  
 901 CAGGCGCAGCGGACGCGCGCGGTGTTGGGTATACCGGTGCTGGCGGTGCTTATCTC 960  
 321 ProArgProAspGlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLys 340  
 961 CGCGCGCGCGGACGCTTCTACCGGACGACATGTATCGCTTGTGGCGGACGCGCAG 1020  
 341 TyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyThrIlePheGlyLeu 360  
 1021 TACGCGACAGTCCGCGGTGATCATCGCGACGACGAGGCGGACCTTGTGGCGTC 1080  
 361 SerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIle 380  
 1081 TCTTTCTTGAAGTGCACACACATGCTCAGGACGCGGTACTTCAAGCAGCTTTTCATC 1140  
 381 HisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGln 400  
 1141 CACGCGCAGGATCGGAGATCGACAGTTGATGGCGCGGTACACCGGACGACATCACCCAG 1200  
 401 GlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSer 420  
 1201 GGTTCCTCCGTCGACACCGGCGATCTTCAATGGCATCACCCTCGGATTTCAACGAGTCTCT 1260  
 421 AlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGln 440  
 1261 GGTTCCTGGCGACCTTGGTTCAGCTTGGCGGTGCGTACTTCTCACTACTTACCAG 1320  
 441 GlyGlyThrLysThrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThr 460  
 1321 GGCGGCGACCAAGTACTCTCTCTAAGCAGCTTCTGGGTGCGGCTCTTCTGGCGAC 1380  
 461 PheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyr 480  
 1381 TTCACGCGCAACGACATCATCTGGCAGGACTACTTGGTGGCGAGCGGAGTGTGATCTAC 1440  
 481 AsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuLeuVal 500  
 1441 AACACGCGTTCATTGGTGTGCGACGACCTCGACCGGACGCGGCGGCTTGTGGAC 1500  
 501 AsnTrpProLysTyrThrSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAla 520  
 1501 AACTGGCCCGACGATACACGAGCTCTCAGTCTGGCAACAACCTTGATGACGATCAACGCG 1560  
 521 LeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAspAlaLeuMetThr 540  
 1561 TTGGGGTTGTACACCGGACGAGCAACTTCGCGCGGATCGGTACGCGGCGCTCTTTTC 1620  
 541 AsnProSerSerPhePheVal 547

Db 1621 AACCCGCCGCTCTTCTTTGTG 1641  
 RESULT 10  
 AAQ54020  
 ID AAQ54020 standard; DNA; 1828 BP.  
 XX AAQ54020;  
 XX 16-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 19-JUL-1994 (first entry)  
 XX Lipase coding sequence of Geotrichum candidum.  
 XX Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;  
 KW Major cell wall protein; glycosyl-phosphatidyl-inositol;  
 KW anchoring protein; alpha factor; alpha-agglutinin; invertase; inulinase;  
 KW alpha-amylase; Geotrichum candidum; lipase; enzymatic process;  
 KW fermentation; biodegradation; catalysis; ss.  
 OS Galactomyces geotrichum.  
 XX FH Key Location/Qualifiers  
 FT CDS 40..1731  
 FT /tag= a  
 FT /product= "Lipase."  
 FT sig\_peptide 40..96  
 FT /tag= b  
 FT mat\_peptide 97..1728  
 FT /tag= c  
 FT /product= "Lipase."  
 XX WO9401567-A1.  
 XX 20-JAN-1994.  
 XX 07-JUL-1993; 93WO-EP001763.  
 XX 08-JUL-1992; 92EP-00202080.  
 XX 14-DEC-1992; 92EP-00203899.  
 XX (UNIL ) UNILEVER PLC.  
 XX (UNIL ) UNILEVER NV.  
 XX Klis FM, Schreuder MP, Toschka H, Verrips CT;  
 DR WPI; 1994-035071/04.  
 DR P-PSDB; AAR47577.  
 XX Immobilisation of enzymes to microbial cell wall - by prodn. of fusion  
 PT protein of enzyme linked to anchoring protein.  
 XX Claim 8; Page 49-52; 99pp; English.  
 CC The lipase is used in a method to immobilise enzymes to a microbial cell  
 CC wall. The coding sequence is used in the production of a recombinant  
 CC polynucleotide which comprises a structural gene encoding a protein with  
 CC catalytic activity (the lipase) and at least part of a gene encoding at  
 CC least the C-terminus of a protein capable of anchoring in a eukaryotic or  
 CC prokaryotic cell wall. The anchoring fragment or protein is selected from  
 CC alpha agglutinin, AGA 1, FLO 1, major cell wall protein of lower  
 CC eukaryotes or a proteinase of lactic acid bacteria. The recombinant  
 CC polynucleotide preferably also comprises a sequence encoding a signal  
 CC peptide to ensure secretion of the expressed product. The signal peptide  
 CC is preferably derived from glycosyl-phosphatidyl-inositol, anchoring  
 CC protein, alpha factor, alpha-agglutinin, invertase or inulinase, alpha-  
 CC amylase of Bacillus or proteinases of lactic acid bacteria. The host  
 CC microorganism can be used for performing enzymatic processes on an  
 CC industrial scale. (Updated on 25-MAR-2003 to correct PN field.) (Updated  
 CC on 16-OCT-2003 to standardise OS field)  
 XX Sequence 1828 BP; 389 A; 483 C; 419 G; 537 T; 0 U; 0 Other;

967 ACCAAGTCCAGGAGTCTTGCACAGTGGCAGAACTCGTATGATCTTTAAGAGCCTGTTT 1026  
 QY 306 -----ThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgPro 323  
 Db 1027 GGTCTCCTCCCTCAATTCCTTGGATTGGT-----CCAGACCC 1065  
 QY 324 AspGlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSer 343  
 Db 1066 GACGCAACATTAITTCGCGATCGCGCTTATGAGCTCTACCGACGGGTAGATACGCCAAG 1125  
 QY 344 ValProValIleIleGlyAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeu 363  
 Db 1126 GTTCCCTCATTTACTGCGCAACACGAGGATGAGGTATATTTCTTGCCCGCTGCTATT 1185  
 QY 364 AsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHisAlaSer 383  
 Db 1186 AATGCTACCACTACTCCCATCTTAAGAAAGTGGTTGAAGTACATTTGTAGCCAGGCTTCT 1245  
 QY 384 AspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerPro 403  
 Db 1246 GACGCTTCGCTTGATCGTGTTTGTGCGCTCTACCCCGGCTCTTGGTCGAGGGGTTCACCA 1305  
 QY 404 PheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeu 423  
 Db 1306 TTCGCACTGGTATCTTTAAATGCTCTTACCCCTCAGTCAAGCGCATTTGTCGCATTTTC 1365  
 QY 424 GlyAspLeuAlaPheIleHisAlaArgTyrPheLeuAsnHisPheGlnGlyGlyThr 443  
 Db 1366 ACTGATTGCTGTTCCAGTCTCCCTGCTGCTGTTATGCTTAACGCTACCAAGGAGCTCAAC 1425  
 QY 444 LysTyrSerPheLeuSerLysGlnLeuSerGlyLeu---ProIleMetGlyThrPheHis 462  
 Db 1426 CGCTGGACTTACCTTGGCACCAGCCAGCTCCATAACCTCGTCCATTTTGGGTACTTTCAT 1485  
 QY 463 AlaAsnAspIleValTrpGlnAspTyrLeu---LeuGlySerGlySerValIleTyrAsn 481  
 Db 1486 GGCAGTGAATCTCTTTTCAATACTACGTGGACCTTGGCCCATCTTCTGCT---TACCGC 1542  
 QY 482 AsnAlaPheIleAlaPheAlaThrAspLeuAsnProAsnThrAlaGlyLeuLeuValAsn 501  
 Db 1543 CGCTACTTTATCTCGTTTGGCAACCACACGACCCCAACCTTGGTATCCAACTCCCAACAG 1602  
 QY 502 TrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAlaLeu 521  
 Db 1603 TGGGATATGTACACT-----GATGCAAGCAAGGAGATGCTTCAGATTTCATATGATT 1653  
 QY 522 GlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGly 534  
 Db 1654 GGTAACCTATGAGAACTGACGACTTTAGAATCGAGGGA 1692  
 RESULT 11  
 AAQ10313  
 ID AAQ10313 standard; DNA; 1692 BP.  
 XX  
 AC AAQ10313;  
 XX  
 DT 24-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 05-APR-1991 (first entry)  
 XX  
 DE Sequence encoding protein with lipase activity.  
 XX  
 KW ATCC 34614; ds.  
 XX  
 OS Galactomyces geotrichum.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1692  
 FT /\*tag= a  
 XX  
 FN JP0229588-A.  
 XX  
 PD 11-DEC-1990.

XX PF 27-MAR-1989; 89JP-00074721.  
 XX PA 27-MAR-1989; 89JP-00074721.  
 XX PA (KURK ) KURITA WATER IND LTD.  
 XX PA (OSAK ) OSAKA CITY.  
 XX DR WPI; 1991-027567/04.  
 XX DR P-PSDB; AAR10330.  
 XX PT Gene for coding protein with lipase activity - is prepd. from messenger  
 XX PT ribonucleic acid of geo-trichum candidum ATCC 34614.  
 XX PS Claim 1; Fig 4; 12pp; Japanese.  
 XX CC The gene product may be isolated from a transformed expression sytem, and  
 CC CC may be enhanced with stability in heat, alkali, acid and organic solvent  
 CC CC by position-specific modulation. (Updated on 25-MAR-2003 to correct PA  
 CC CC field.) (Updated on 24-OCT-2003 to standardise OS field)  
 XX SQ Sequence 1692 BP; 336 A; 470 C; 399 G; 487 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 1-26e-94 Length: 1692  
 Score: 1097.00 Matches: 238  
 Percent Similarity: 56.16% Conservatives: 72  
 Best Local Similarity: 43.12% Mismatches: 200  
 Query Match: 38.42% Indels: 42  
 DB: 2 Gaps: 8

US-09-943-857-4 (1-547) x AAQ10313 (1-1692)

QY 9 GlyArgLeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeu 28  
 DB 46 GGCACCTTGGCCAGCCGCCCGCTTCTTAATGGCAACGAGGTCTCTGTGTGC 105  
 QY 29 AsnAlaIleAsnGluAlaPheLeuGlyIleProPheAlaGluProValGlyAsn 48  
 DB 106 CTTGAGGGCAAGGTTGATCTCAAGGAAATCCCAATTTGCTGACCTCTGTGTGTAC 165  
 QY 49 LeuArgPheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSer 68  
 DB 166 TTGCGGTTCAAGCACCACCCAGCCTTTCACTGGATCTTACCAGGCTTTAAGGCCAACGAC 225  
 QY 69 TyrGlyProSerCysMetGlnGlnAsnProGluGlyThrPheGlu----- 83  
 DB 226 TTCAAGCTCTGTGTATGACGCTTGATCTCTGGCAATGCCATTTCTTTGTCTTGACAAAGTC 285  
 QY 84 -----GluAsnLeuGlyLysThrAlaLeuAspLeuValMet 95  
 DB 286 GTGGGCTTGGGAAAGATTATCTCTGATACCTTAGAGGCCCTCTTTATGACATGGCC--- 342  
 QY 96 GlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnVal 115  
 DB 343 -----CAGGCTAGTGTTCATCAATGAGGACGTCTCTACCTTAACGTT 387  
 QY 116 ValArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuThrIlePheGly 135  
 DB 388 TTCCGCGCCGCTGGCCCAAGCCTGATCTAAGCTCCCGCTCATGGTTGATTTACGGT 447  
 QY 136 GlyGlyPheGluIleGlySerProThrIlePheProAlaGlnMetValThrLysSer 155  
 DB 448 GGTGCTCTTGTGTGTGTCTTCTGCTTCTTACCTGGTAAACGCTTACGTTCAAGGAGAGT 507  
 QY 156 ValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrpGly 175  
 DB 508 CTGGAATGGCCAGCCTGT 567  
 QY 176 PheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGln 195  
 DB 568 TTCCTGGGTGTGTATGCCATCACCGCTGAGGGTAACCAACCAACGCTGTGTCTGTGACGACCAG 627

QY 196 ArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLys 215  
 DB 628 CGCAAGGCTCTCGAGTGGTGTAGCAACAACATTCGCACTTTGGTGGTGTATCCGACAG 687  
 QY 216 ValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsn 235  
 DB 688 GTCATGATTTTTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 747  
 QY 236 AspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGly 255  
 DB 748 GGTGGTGAACAACATCAACGGAAGCAGCTTTTCCACTCTGCCATTTCTTCAGTCTGGC 807  
 QY 256 AlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPhe 275  
 DB 808 GGTCTCTCTCTTACTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 867  
 QY 276 ValSerSerAlaGlyCys--GlySerAlaSerAspLys-----LeuAlaCysLeuArg 292  
 DB 868 GCTCATGATGCGGATGTGATGCCAGCCCGGTGCAATGAACACTCTGCTTGTCTCCGC 927  
 QY 293 SerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn----- 305  
 DB 928 AGCAAGTCCAGCGATGCTTGCACAGTGCACAGACTCGTACGATCTCAAGGACCTGTTT 987  
 QY 306 -----ThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgPro 323  
 DB 988 GGTCTGCTCCCTCAATTCCTTGGATTGGT-----CCGAGACCC 1026  
 QY 324 AspGlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSer 343  
 DB 1027 GACGGCAACATTATTCCTGATGCCCTTATGAACTCTACCGCAGCGGTAGATACGCCAAG 1086  
 QY 344 ValProValIleIleGlyAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeu 363  
 DB 1087 GTTCTCTACATTTACTGTAAACGAGGATGAGGTACTATTCTTCCGCCCTGGGTGTTT 1146  
 QY 364 AsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHisAlaSer 383  
 DB 1147 AATGTACACACTTCTCCCATGTTAAGAGTGTGTAAGTACATTTTGTAGCAGGCTTCT 1206  
 QY 384 AspAlaGluIleAspThrLeuMetAlaTyrProGlnAspIleThrGlnGlySerPro 403  
 DB 1207 GACGCTTGGTGTGCTGCTTGTGCTTACCCCGGCTCTTGTGCGAGGGTGGCCA 1266  
 QY 404 PheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeu 423  
 DB 1267 TTCGCACTGGCATCTCAATGCTCTGACCCCTCAATTCAAGCGCATTTGCTGCAATTTTC 1326  
 QY 424 GlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGlnGlyGlyThr 443  
 DB 1327 ACTGATTTTCTGTTCCAGTCTCTCTGCTGCTTATGCTTAAAGCTACCAAGGACGTCAC 1386  
 QY 444 LysTyrSerPheLeuSerLysGlnLeuSerGlyLeu---ProIleMetGlyThrPheHis 462  
 DB 1387 CGCTGAGTACTTCTTCCACCCAGCTCCATACCTGTTCCATTTCTTGGTACTTTCCAT 1446  
 QY 463 AlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyrAsnAsn 482  
 DB 1447 GGTAGTACTTCTTCTTCCAAATACCTACGCTGGACCTTGGTCTCTCTGTTACCCCGC 1506  
 QY 483 AlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuValAsnTrp 502  
 DB 1507 TACTTTATCTGTTTCCCAACCCAGCAGCCCAACGTTGCTGCAACCACTGAAACAGTGG 1566  
 QY 503 ProLysTyrThrSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAlaLeuGly 522  
 DB 1567 GATATGTACT-----GATTTCAGCAAGGAGATGTTTCAGATTCATATGATTGGT 1617  
 QY 523 LeuTyrThrGlyLysAspAsnPheArgThrAlaGly 534  
 DB 1618 AACTTATGAGAACTGACGACTTTAGAAATCGGGA 1653

AAT10421  
 ID AAT10421 standard; DNA; 1635 BP.  
 AC  
 AAT10421;  
 XX  
 DT 16-OCT-2003 (revised)  
 DT 20-SEP-1996 (first entry)  
 XX  
 DE Geotrichum candidum lipase gene.  
 XX  
 DE  
 XX  
 KW Fatty acid; transgenic plant; exogenous; lipase; Rhizopus niveus;  
 KW Pseudomonas aeruginosa; Pseudomonas fluorescens; Geotrichum candidum;  
 KW Candida cylindracea; constitutive; tissue specific; promoter; lipid;  
 KW milling; biofuel; lubricant; detergent; ss.  
 XX  
 OS Galactomyces geotrichum.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 CDS 1..1635  
 FT /\*tag= a  
 FT /product= "lipase protein"  
 XX  
 XX  
 PN FR2722798-A1.  
 XX  
 XX  
 PD 26-JAN-1996.  
 XX  
 PF 25-JUL-1994; 94FR-00009272.  
 XX  
 PR 25-JUL-1994; 94FR-00009272.  
 XX  
 PA (NAPO-) INST NAT POLYTECHNIQUE TOULOUSE.  
 XX  
 PI Alibert G, Mouloungui Z, Boudet A;  
 DR WPI; 1996-107680/12.  
 XX  
 PT Prodn. of fatty acids or derivs. from transgenic oilseed plants -  
 PT engineered to express a lipase that contacts lipid(s) only when seeds are  
 PT milled.  
 XX  
 PS Claim 9; Page 25-26; 32pp; French.  
 XX  
 CC A novel method of producing fatty acids or their derivs. in a plant  
 CC comprises generating a transgenic plant contg. an exogenous lipase gene,  
 CC esp. selected from the Rhizopus niveus, Pseudomonas aeruginosa,  
 CC P.fluorescens, Pseudomonas sp., Geotrichum candidum or Candida  
 CC cylindracea lipase genes (AAT10417-22 resp.). The lipase genes can be  
 CC placed under control of a constitutive or a tissue specific promoter. The  
 CC production of the fatty acids only occurs when the lipase and lipids  
 CC contact each other after milling of the plants. The fatty acids generated  
 CC can be used to prod. e.g. biofuels, lubricants, detergents, etc. (Updated  
 CC on 16-OCT-2003 to standardise OS field)  
 XX  
 SQ Sequence 1635 BP; 323 A; 448 C; 395 G; 469 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 8-7e-94 Length: 1635  
 Score: 1098.00 Matches: 238  
 Percent Similarity: 56.86% Conservative: 73  
 Best Local Similarity: 43.51% Mismatches: 192  
 Query Match: 37.99% Indels: 44  
 DB: 2 Gaps: 10

US-09-943-857-4 (1-547) x AAT10421 (1-1635)

QY 15 ProThAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleLeuAsnGlu 34  
 DB  
 DB 7 CCCACGCCGCTTCTAATGCAACGAGGTCTCTGGTGTCTCTGGGGCAAGGTTGAT 66  
 QY 35 AlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro 54  
 DB  
 DB 67 ACCTTTAAGGGAATCCATTGTGTGACCTCTCTGGTGTGACTTGGGTTCAAGCACCCC 126

QY 55 ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMet 74  
 DB  
 DB 127 CAGCCTTTCACTGGATCCTACCAAGGCTTTAAGGCCCAACGACTTCAGCTCTGCTTGTATG 186  
 QY 75 GlnGlnAsnProGluGlyThrPheGlu----- 83  
 DB  
 DB 187 CAGCTTGTATCTGGCAATGCCATTTCTTGGCTTGCACAAAGTCGTGGGCTTGGGAAAGATT 246  
 QY 84 -----GluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101  
 DB  
 DB 247 CTTCTGTGATAACCTTAGAGCCCTCTTTATGATGAGCC-----CAG 288  
 QY 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgProProGlyThr 121  
 DB  
 DB 289 GGTAGTGTCTCCATGAATGAGGACTGCTCTACCTTAACGTTTCCGCCCTTGTGGCACC 348  
 QY 122 LysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluLeuGly 141  
 DB  
 DB 349 AAGCCTGATGCTAAGCTCCCGCTCATGTTTGGATTACGGTGGTGCCTTTGTGTTGGT 408  
 QY 142 SerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLysHis 161  
 DB  
 DB 409 TCTTCTGCTTCTTACCTGGTAACGGCTAGCTCAAGAGAGAGTGTGAAATGGGCCAGCCT 468  
 QY 162 IleIleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAspAsp 181  
 DB  
 DB 469 GTTGTGTTGTTTCCATCACTACCGTACCGGCCCTCATGATTCCTGGTGGTATGCC 528  
 QY 182 IleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrp 201  
 DB  
 DB 529 ATCACCGCTGAGGGTAACCAACCGTGTCTGCAGCACCCAGCGCAAGGGTCTCGAGTGG 588  
 QY 202 ValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePheGlyGlu 221  
 DB  
 DB 589 GTTAGCGACAACCAATTGCCCACTTTGGTGGTATCCCGACAAGGTATGATTTTCGGTGAG 648  
 QY 222 SerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsnThrTyr 241  
 DB  
 DB 649 TCGCTGTGTGCCATGAGTGTGCTCACAGCTTGTTCCTACCGTGGTGGTGCACACACTAC 708  
 QY 242 LysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSerAsp 261  
 DB  
 DB 709 AACGGAAGAAAGCTTTTCCACTCTGCATCTCTCAGTCTGGCGGTCTCTCTCTACTTT 768  
 QY 262 ProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGlyCys 281  
 DB  
 DB 769 GACTCTACTTCTGTGGTCCGAGAGTGGCTACAGCAGATTGCTCAGTATGCCGATGT 828  
 QY 282 ---GlySerAlaSerAspLys-----LeuAlaCysLeuArgSerAlaSerAspThr 298  
 DB  
 DB 829 GATGCCAGCGCCAGTGCACATGAACCTCTGGCTTGTCTCCGACAGAGTCCAGCGATGTC 888  
 QY 299 LeuLeuAspAlaThrAsnAsn-----ThrProGlyPhe 309  
 DB  
 DB 889 TTGCACAGTGCACCAAGCTCTCAGCATCTCAAGGACCTGTTGGCCCTGCTCCCTCAATTC 948  
 QY 310 LeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThr 329  
 DB  
 DB 949 CTTGGATTGGT-----CCAGACCCCGACGCAACATTATTTC 987  
 QY 330 AspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGly 349  
 DB  
 DB 988 GATCCGCTTATGAGCTCTACCGCAGCGGTAGATAGCCCAAGGTTCCCTACATTACTGTT 1047  
 QY 350 AspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAla 369  
 DB  
 DB 1048 AACCCAGGAGTATGAGGTTACTATTCTTCCCGCTGCTGTTAATGCTACCAGACTCC 1107  
 QY 370 GlnAlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluLeuAspThr 389  
 DB  
 DB 1108 CATGTTAAGAGTGGTGAAGTACATTTGTAGCAGGCTTCTGACGCTTCGCTGTGATCGT 1167  
 QY 390 LeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeu 409

```

Db 1168 GTTTGTGCTCTACCCCGGCTCTTGTGCGAGGCTGCGCATTCGCACTGATCTT 1227
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 410 AsnAlaLeuThrProGlnPheLeuArgIleSerAlaValLeuGlyAspLeuAlaPheIle 429
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1228 AATGCTCTGACCCCTCAGTTCAGGCGCATTCGCGCATTTTTCACCTGATTTGCTGTCAG 1287
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 430 HisAlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSer 449
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1288 TCTCCTCGTGTGTTATGCTTAACGCTACCAAGGAGCGTCAACCGCTGGACTTACCTGGCC 1347
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 450 LysGlnLeuSerGlyLeu---ProIleMetGlyThrPheHisAlaAsnAspIleValTrp 468
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1348 ACCAGCTCCATAACCTCGTTCATTTTGGGTACTTTCCATGGTGTAGTGAICTCTTTTC 1407
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 469 GlnAspTyrLeu---LeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPhe 487
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1408 CAATACTACGTGGACCTTGGCCCATCTTCTGCT---TACCCTCGCTACTTTTATCTCGTT 1464
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 488 AlaThrAspLeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSer 507
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1465 GCCAACCCACGACCCCAAGCTGGGACCAACCTGAACAGTGGGATATGTACACT--- 1521
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 508 SerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLys 527
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1522 -----GATGAGGCAAGGAGATGCTTCAGATTTCATATGTTGGTAACTCTATGAGAACT 1575
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 528 AspAsnPheArgThrAlaGly 534
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1576 GACGACTTTAGAAATCGAGGGA 1596
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

## RESULT 13

AAQ05605 standard; cDNA; 1674 BP.

XX AAQ05605;

XX 24-OCT-2003 (revised)

DT 25-MAR-2003 (revised)

DT 13-DEC-1990 (first entry)

XX Gene encoding protein with lipase activity.

XX Lipase.

XX Galactomyces geotrichum; ATCC34614.

XX Key Location/Qualifiers

FT misc\_RNA 40..1671

FT /tag= a

FT /label= Claim 1

XX JP02174680-A.

XX 06-JUL-1990.

XX 27-DEC-1988; 88JP-00330598.

XX 27-DEC-1988; 88JP-00330598.

XX (KURITA) KURITA WATER IND LTD.

XX (OSAKA) OSAKA CITY.

XX WPI; 1990-250686/33.

XX P-PSDB; AAR96370.

XX Protein gene having lipase activity - has defined sequence of 544 bases  
 PT and gives lipase producing vector on integration to expression  
 PT (secretion) vector.

XX Disclosure; Fig 4; 12pp; Japanese.

XX The cloned cDNA can be inserted into an expression vector and used to

CC transform hosts for the prodn. of a protein with lipase activity. See  
 CC also AAQ05605. (Updated on 25-MAR-2003 to correct PA field.) (Updated on  
 CC 24-OCT-2003 to standardise OS field)

XX Sequence 1674 BP; 351 A; 473 C; 381 G; 469 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 8.08e-93 Length: 1674  
 Score: 1078.00 Matches: 233  
 Percent Similarity: 56.43% Conservative: 83  
 Best Local Similarity: 41.61% Mismatches: 202  
 Query Match: 37.64% Indels: 42  
 DB: 2 Gaps: 8

US-09-943-857-4 (1-547) x AAQ05605 (1-1674)

```

Qy 8 AlaGlyArgLeuGlySerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGly 27
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 25 GCTGGCGTCTTGGCCGAGCCGCGGCTCTCTCAATGGCAAGAGGTCTCTCTGCT 84
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 28 LeuAsnAlaIleIleAsnGluAlaPheLeuGlyIleProPheAlaGluProProValGly 47
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 85 GTCCTTGAGGCAAGGTGTATACCTTCAAGGGAATCCATTTGCTGACCCCTCTTGAAT 144
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 48 AsnLeuArgPheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThr 67
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 145 GACTTGGATTCAAGCACCCCGAGCTTTCATGATCTCTACAGGCTCTTAAAGCCCAAT 204
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 68 SerTyrGlyProSerCysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGly 87
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 205 GATTTTAGCCCTGCTTGTATGACGCTTGATCCTGGCAACTCTCTC----- 249
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 88 LysThrAlaLeuAspLeuValMet---GlnSerLysValPhe----- 100
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 250 ---ACTTTGCTTGACAAAGCTCTGGGATTGCGAAAGTCAATCCCGAAGAAATTTAGAGT 306
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 101 -----GlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsn 114
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 307 CCCCTTTATGATATATGCGGCAAGGTACCGTGTGATGATGAGGACTGTCTTTACCTCAAT 366
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 115 ValValArgProProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePhe 134
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 367 GTTTTCGCGCTGCTGGCACCAGCTGATGCTAAGCTCCCGCTCATGCTTGTGGATTATAC 426
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 135 GlyGlyGlyPheGluIleGlySerProThrIlePheProProAlaGlnMetValThrLys 154
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 427 GGTGTGCTTGTGTTTACGTTCTCTGCTGCTTACCTGGTAAACAGCTACGTTAAGGAA 486
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 155 SerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrp 174
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 487 AGTATCAACATGGGCGAGCCGCTGTGTGTTTCCATCACTACCGTACCGGTCCATTT 546
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 175 GlyPheLeuAlaGlyAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAsp 194
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 547 GGATTCTGCGGTGGTGTGCCATCACCGCTGAGGCGAACCAACGCTGGTGTGACGAC 606
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 195 GlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyAspProSer 214
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 607 CAGCGCAAGGCTCTCGAGTGGGTAGCGCAACATTCGCAACTTTGGTGTGATCCCGAT 666
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 215 LysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrp 234
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 667 AAGGTCAATGATTTTCGGTGAAGTCCCGTGTGCTGAGTGTGCTCACAGCTTATGTC 726
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 235 AsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSer 254
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 727 TATGTTGTGACCAACCTTACCAACGAAAGAGCTTTTCCACTTGGCTTCTTCTCAGTCT 786
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 255 GlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeu 274
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 787 GGTGGCCCTCTTCTTACCACGACTCTAGTCCGTTGGTCCGATATTTCTTACACAGA 846
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 275 PheValSerSerAlaGlyCys-----GlySerAlaSerAspLysLeuAlaCysLeu 291
      :::::::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

```

Db      847  TTTGCTCAGTATCGGATGTGACACTAGTCCAGTCCCAACGACACTCTGGAGTGTCTC 906
QY      292  ArgSerAlaSerSerSerThrLeuLeuAspAlaThrAsnAsn----- 305
Db      907  CGCAGCAAGTCCAGCTGCTCTTGACAGATGCCAGAACTCTTACGATCTCAAGATCTG 966
QY      306  -----ThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArg 322
Db      967  TTTGGTCTACTCCCTCAATTCCTTGGATTGGT-----CCAGA 1005
QY      323  ProAspGlyLysAsnLeuThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAla 342
Db      1006  CCCAGCGGCAACATTATTCGCGATGCCGCTTATGAGCTCTTCGCGACGGTAGATACGCC 1065
QY      343  SerValProValIleLeuGlyAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSer 362
Db      1066  AAGGTTCCCTACATTAGCGGTAAACCGAGAGATGAAGTACTGCTTTTGTCTCTGTGTG 1125
QY      363  LeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHisAla 382
Db      1126  CTCACGCTACCAAGCTCCCATGTTAAGAGTGGTTGCAGTACATTTTCTACGATGCT 1185
QY      383  SerAspAlaGluIleAspThrLeuMetAlaTyrProGlnAspIleThrGlnGlySer 402
Db      1186  TCCGAGGCTTCCATTGACCGTGTGTTGCTGCTGTACCCGCGACACCTCTCTGTGTGCTG 1245
QY      403  ProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaVal 422
Db      1246  CCTTCGCACTGGCATCTTAATGCGCTACCCCGCTCAAGCGTGTGGCGGCATC 1305
QY      423  LeuGlyAspLeuAlaPheIleHisAlaArgTyrPheLeuAsnHisPheGlnGlyGly 442
Db      1306  TTGTCGATATGCTTTTCCAGTCTCCCGCGCGTGATGCTTAGCGCCACCAAGGAGTT 1365
QY      443  ThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeu---ProIleMetGlyThrPhe 461
Db      1366  AACCGCTGACATTACCTTTCCGCCCATCTGCACAACCTGCTGCCAATTTTGGGTACTTTC 1425
QY      462  HisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyrAsn 481
Db      1426  CATGGCAACGAGCTTATCTTCCAAATCAATGTAACATTCGCGCCCGCTAACTCCTACCTT 1485
QY      482  AsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuLeuValAsn 501
Db      1486  CGTTATTTTATTTCCTTTGCCACCAACCATGACCTTAATGTTGGTACTAATCTGCTCCAG 1545
QY      502  TrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAlaLeu 521
Db      1546  TGGGATCAATACACT-----GATGAAGGCAAGGAGATGCTTGAGATTACATGACC 1596
QY      522  GlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAspAlaLeuMetThrAsn 541
Db      1597  GATAATGTCATGAGACTGATGACTACAGAAATTGAGGGAATCTCAAACTTTGAGACTGAC 1656

RESULT 14
ID      AAQ98578
XX      AAQ98578 standard; DNA; 2045 BP.
AC      AAQ98578;
XX
DT      16-OCT-2003 (revised)
DT      05-MAR-1996 (first entry)
XX
DE      Aspergillus foetidus glucoamylase gene promoter.
XX
KW      recombinant; glucose oxidase; GOD; extracellular production;
KW      filamentous fungus; glucoamylase; promoter; ss.
XX
OS      Aspergillus foetidus; SE4.
XX
PN      EP65291-A1.
XX

```

```

PD      02-AUG-1995.
XX
PF      19-JAN-1995; 95EP-00200129.
XX
PR      28-JAN-1994; 94BE-00000102.
PR      17-JUN-1994; 94BE-00000586.
PR      09-JAN-1995; 95BE-00000014.
XX
PA      (SOLV ) SOLVAY & CIE.
XX
PI      Carrez D, Roos J;
XX
DR      WPI; 1995-264864/35.
XX
PT      Expression system for extracellular prodn. of glucose oxidase - related
PT      vectors and transformed cells, esp. new strain of Aspergillus foetidus,
PT      provides high yields of enzyme uncontaminated by catalase.
XX
PS      Claim 18; Page 17-19; 32pp; French.
XX
SS      The promoter and terminator sequences from the glucoamylase gene of
CC      Aspergillus foetidus SE4 strain (AAQ98578 and AAQ98570, respectively) are
CC      used for extracellular expression of glucose oxidase in A.foetidus hosts.
CC      The specifically claimed transformant strain SE4tr contains a vector in
CC      which the GOD gene is under transcriptional control of the new promoter
CC      and terminator sequences; extracellular production of GOD by the
CC      transformant was 5000-10000 times greater than for the untransformed
CC      parental SE4 strain. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ      Sequence 2045 BP; 454 A; 538 C; 565 G; 488 T; 0 U; 0 Other;

```

```

Alignment Scores:
Pred. No.: 2,16e-79 Length: 2045
Score: 938.50 Matches: 207
Percent Similarity: 60.81% Conservative: 80
Best Local Similarity: 43.86% Mismatches: 162
Query Match: 32.77% Indels: 24
DB: 2 Gaps: 8

US-09-943-857-4 (1-547) x AAQ98578 (1-2045)
QY      89 ThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGlu 108
Db      13 TCGGTATTGGCTCTCTCGCTGATCTCCCTTTGGTACAGTCGCTACCAATGCTCTCGAG 72
QY      109 AspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAlaAsnLeuPro 128
Db      73 GATTGCTGAACATTGACATTCCGCGTCCGCGCGGACCAACCCGCGACTCGAAGCTGCCT 132
QY      129 ValMetLeuTriPhePheGlyGlyPheGluIleGlySerProThrIlePheProPro 148
Db      133 GTGCTGGTCTGGATCTTTGGCGGAGCTTTGAACCTTGGTTCAAGCGGATGATGATGTT 192
QY      149 AlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsn 168
Db      193 ACAACGATGGTATCATCGCTGATAGACAAGAACATCGCTATCGTTTGTAGCAATGAAT 252
QY      169 TyrArgValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGly 188
Db      253 TATCGCGTGGGAGGTTTCGGGTTCTTCCCGGAAGGAGATCCTCGAGGACGGGTCGCG 312
QY      189 AsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGly 208
Db      313 AACCTAGGGCTCTCTGGACCAACGCTTGGCTTGCAGTGGGTTGCCACAAACATCGAGGCC 372
QY      209 PheGlyGlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerVal 228
Db      373 TTTGGTGGAGACCCGAGCAAGGTGACGATTGGGGAGAAATCAGCAGGAGCCCATTCGTTTG 432
QY      229 -LeuCysHisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheAr 248
Db      433 ACTAGATGACTTGTGAC-----GACGGAAACATCACTTACAAGGATAAGCCCTTGTTCGG 486

```

QY 248 gAlaGlyIleMetGlnSerGlyValaMetValProSerAspProValAspGlyThrTyrgl 268  
 Db 487 GGGGGCCATCATGACTCCGCTAGTGTGTTCCCGCAGACCCCGTCGATGGGTCAAGGG 546  
 QY 268 yAsnGluIleTyrrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLe 288  
 Db 547 ACAGCAAGTATATGATGGGTAGTGAATCTGCAGGCTGTCTCTTCTTAACGACACCC 606  
 QY 288 uAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnThrProgl 308  
 Db 607 AGCTGTGCTGGTGAACCTAGACTACACCGACTTCCTCAATGGCGAAACTCCGTGCCAGG 666  
 QY 308 yPheLeuAlaTyrrSerSerLeuArgLeuSerTyrrLeuProArgProAspGlyLysAsnIl 328  
 Db 667 CATTTAAGTACCATCTCTGGCGTTATCATATGTGCTCGACCGGACGGCGCGCT 726  
 QY 328 eThrAspAspMetTyrrLysLeuValArgAspGlyLysTyrrAlaSerValProValIleIl 348  
 Db 727 GTCGGCATCACCGGACGTTTGGGCAAGACGAGGAATATGTCGGGTCCCGTTTCATCGT 786  
 QY 348 eGlyAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAs 368  
 Db 787 GGGGACACAGAGGATAGGGACCTTATTCGCTTGTTCAGTCCAACTTACGACCAT 846  
 QY 368 nAlaGlnAlaArgAlaTyrrPheLysGlnSerPheIleHis---AlaSerAspAlaGluIl 387  
 Db 847 CGACGAGTGTGCTGACTACCTGGCCTCATCTTCTTATGACGCTAGCCGAGAGCAGCT 906  
 QY 387 eAspThrLeuMetAlaAlaTyrrProGlnAspIleThrGlnGlySerProPhe-AspThrG 407  
 Db 907 TGAAGAACTAGTGGCCCTGTATCCACAGACACCACGCTACGCGGTCTCGGTTTCAGACGG 966  
 QY 407 lyValLeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuA 427  
 Db 967 CGGCC---AACACTGGTATCCGCAATTGAAGCAATTGGCCCACTTCTCGCGCACTGG 1023  
 QY 427 laPheIleHisAlaArgArgTyrrPheLeuAsnHisPheGlnGlyGlyThr----- 443  
 Db 1024 TCTTCAC-CATTAACCGGGGCACTTCTCTGATGCAGAGGAATCTCCCTGATCTTC 1082  
 QY 444 --LysTyrrSerPheLeuSerGlnLeuSerGlyLeuProIleMetGlyThrPheHisA 463  
 Db 1083 CGAACTGGTGTGCTGACCTGGCGACCTATGACTATGGCACCCTTCTCTGGGACCTTCCACG 1142  
 QY 463 laAsnAspIleValTrpGlnAspTyrr-----LeuLeuGlySerGlySerV 478  
 Db 1143 GAAGTGACCTGCTGACGGTGTCTATGGATCGACGCCAACTATCGAGTAGTCTTAGCC 1202  
 QY 478 alIleTyrrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThr---AlaG 497  
 Db 1203 ACACGTAC-----TATCTGAGCTTTGTGTATAGCTGATCGGAATCCGAACCCCGG 1253  
 QY 497 lyLeuLeuValAsnTrpProLysTyrrThrSerSerSerGlnSerGlyAsnAsnLeuMetM 517  
 Db 1254 GGGAGTACATTAGTGGCGCGCAGTGGGAAGGAATCGCGGCGAG-----TTGATGA 1301  
 QY 517 eIleAsnAlaLeuGlyLeuTyrrThrGlyLysAspAsnPheArgThrAlaGlyTyrrAspA 537  
 Db 1302 ATTTCGGAGCGAACGACGCGCAGCTCTCTTACGATGATTTCCGCAACCGGACATATGAT 1361  
 QY 537 laLeuMetThrAsnProSerSerPhePheVal 547  
 Db 1362 TCATCTCGCAGAAATACCGCGGCTTCCACATC 1393  
 RESULT 15  
 ID AAT11299  
 XX AAT11299 standard; DNA; 2045 BP.  
 AC AAT11299;  
 XX  
 XX  
 DT 16-OCT-2003 (revised)  
 DT 25-JUL-1996 (first entry)  
 XX

DE Aspergillus foetidus SE4 glucoamylase promoter sequence.  
 XX  
 KW Microbial aspartic protease; filamentous fungus; aspergillopepsin;  
 extracellular production; secretion; glucoamylase gene promoter; ss.  
 XX  
 OS Aspergillus foetidus; SE4.  
 XX  
 PN EP687734-A1.  
 XX  
 PD 20-DEC-1995.  
 XX  
 PF 09-JUN-1995; 95EP-00201521.  
 XX  
 PR 17-JUN-1994; 94BE-00000585.  
 PR 09-JAN-1995; 95BE-00000015.  
 XX  
 PA (SOLV ) SOLVAY SA.  
 XX  
 PI Carrez D, Dhaese P;  
 XX  
 DR WPI; 1996-031780/04.  
 XX  
 PT System for extracellular prodn. of microbial aspartic protease - contg.  
 PT promoter, signal sequence, pro-peptide and mature enzyme sequences, and  
 PT terminator, esp. for aspergillopepsin prodn.  
 XX  
 PS Claim 19; Page 18-20; 35pp; French.  
 XX  
 CC Spores of A. foetidus ATCC 14916 were exposed to UV, then grown on a  
 nutrient medium and one mutant having increased production of  
 CC glucoamylase was selected and designated SE4. The promoter and terminator  
 CC regions of the glucoamylase gene from mutant strain SE4 were isolated and  
 CC sequenced. These transcription control regions are preferred for  
 CC incorporation into novel expression vectors designed to allow  
 CC extracellular production of microbial aspartic protease. In particular,  
 CC the enzyme aspergillopepsin is produced by transformed filamentous fungi  
 CC and secreted into the culture medium due to the presence of a fungal  
 CC secretion peptide. The present sequence is that of the A. foetidus SE4  
 CC glucoamylase promoter. (Updated on 16-OCT-2003 to standardise OS field)  
 XX  
 SQ Sequence 2045 BP; 454 A; 538 C; 565 G; 488 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 2,16e-79 Length: 2045  
 Score: 938.50 Matches: 207  
 Percent Similarity: 60.81% Conservative: 80  
 Best Local Similarity: 43.86% Mismatches: 162  
 Query Match: 32.77% Indels: 24  
 DB: 2 Gaps: 8  
 US-09-943-857-4 (1-547) x AAT11299 (1-2045)  
 QY 89 ThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGlu 108  
 Db 13 TCGGTGATGGCTCTCTCGTGTATCTCCTTTGTGTACAGTCGGTACCATCTCTCGAG 72  
 QY 109 AspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAlaAsnLeuPro 128  
 Db 73 GATTGCTGAACATTGACATTGGCGTCCGGCGGAGCCACCGCGAGCTCGAAGTGCCT 132  
 QY 129 ValMetLeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePheProPro 148  
 Db 133 GTGCTGGTCTGGATCTTTGGCGGAGGCTTGAACCTTGGTTCAAAGCGCATGTATGATGT 192  
 QY 149 AlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsn 168  
 Db 193 ACAACGATGTATCATCTGTGTATACAGACATGCGCTATCGTCTTTGTAGCAATGAT 252  
 QY 169 TyrArgValAlaSerTrpGlyPheLeuAlaGlyAspIleLysAlaGluGlySerGly 188  
 Db 253 TATCGCGTGGGAGGTTTCGGGTTCCTGCCGGAAGAGATCTCTGGAGGACGGTCCGGC 312  
 QY 189 AsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGly 208





B/GNK

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 5, 2004, 19:05:41 ; Search time 91 Seconds  
(without alignments)

3335.804 Million cell updates/sec

Title: US-09-943-857-4

Perfect score: 2864

Sequence: 1 SMNSRGPAGRLGVSPTAKLA.....DNFRTAGYDALMTNPFSSFFV 547

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DB=xlh

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US09943857/runat\_29072004\_090901\_29773/app\_query.fasta\_1.711  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rnl -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09943857 @CGN 1.1 54 @runat\_29072004\_090901\_29773 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_1/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTCUTS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2415	84.3	1650	2	US-08-776-210-6
2	1107	38.7	1928	3	US-08-362-525-11
3	1089	38.0	1635	2	US-08-776-210-5
4	938.5	32.8	2045	1	US-08-379-928A-7
5	632	22.1	8533	1	US-07-846-181-6
6	632	22.1	8533	1	US-07-845-989-6
7	452.5	15.8	1738	2	US-08-932-376A-1
8	452	15.8	1845	1	US-07-732-962A-1
9	452	15.8	1845	5	PCT-US92-06106-1
10	452	15.8	2256	2	US-08-318-826A-5
11	452	15.8	2256	2	US-08-370-156-1
12	452	15.8	2256	3	US-08-814-095-1

13	452	15.8	2400	6	5215909-13	Patent No. 5215909
14	452	15.8	3016	2	US-08-318-826A-7	Sequence 7, Appli
15	452	15.8	3016	2	US-08-370-156-5	Sequence 5, Appli
16	452	15.8	3016	3	US-08-814-095-5	Sequence 5, Appli
17	452	15.8	3096	2	US-08-318-826A-6	Sequence 6, Appli
18	452	15.8	3096	2	US-08-370-156-3	Sequence 3, Appli
19	452	15.8	3096	3	US-08-814-095-3	Sequence 3, Appli
20	451	15.7	2381	2	US-08-318-826A-9	Sequence 9, Appli
21	451	15.7	2416	2	US-08-318-826A-8	Sequence 8, Appli
22	451	15.7	2416	3	US-09-334-489-1	Sequence 1, Appli
23	451	15.7	2416	3	US-09-334-489-2	Sequence 2, Appli
24	444.5	15.5	2445	6	5215909-9	Patent No. 5215909
25	437.5	15.3	1470	1	US-08-589-893-17	Sequence 17, Appl
26	437.5	15.3	1470	2	US-09-020-991-17	Sequence 17, Appl
27	437.5	15.3	1470	2	US-09-062-890-17	Sequence 17, Appl
28	436.5	15.2	1470	1	US-08-589-893-11	Sequence 11, Appl
29	436.5	15.2	1470	2	US-09-062-890-11	Sequence 11, Appl
30	436.5	15.2	1470	2	US-09-062-890-11	Sequence 11, Appl
31	432.5	15.1	1470	1	US-08-589-893-3	Sequence 3, Appli
32	432.5	15.1	1470	1	US-08-589-893-5	Sequence 5, Appli
33	432.5	15.1	1470	1	US-08-589-893-19	Sequence 19, Appl
34	432.5	15.1	1470	2	US-09-020-991-3	Sequence 3, Appli
35	432.5	15.1	1470	2	US-09-020-991-5	Sequence 5, Appli
36	432.5	15.1	1470	2	US-09-020-991-19	Sequence 19, Appl
37	432.5	15.1	1470	2	US-09-062-890-3	Sequence 3, Appli
38	432.5	15.1	1470	2	US-09-062-890-5	Sequence 5, Appli
39	432.5	15.1	1470	2	US-09-062-890-19	Sequence 19, Appl
40	432.5	15.1	1470	2	US-09-062-890-27	Sequence 27, Appl
41	431.5	15.1	1470	1	US-08-124-674-1	Sequence 1, Appli
42	431.5	15.1	1470	1	US-08-589-893-1	Sequence 1, Appli
43	431.5	15.1	1470	2	US-09-020-991-1	Sequence 1, Appli
44	431.5	15.1	1470	2	US-09-062-890-1	Sequence 1, Appli
45	430.5	15.0	1470	1	US-08-589-893-7	Sequence 7, Appli

#### ALIGNMENTS

#### RESULT 1

US-08-776-210-6  
; Sequence 6, Application US/08776210  
; Patent No. 5942659  
; GENERAL INFORMATION:  
; APPLICANT: ALIBERT, Gilbert  
; APPLICANT: MOULOUNGUI, Zephirin  
; APPLICANT: BOUDET, Alain  
; TITLE OF INVENTION: PROCESS FOR PRODUCING FATTY ACIDS OR  
; DERIVATIVES THEREOF FROM OLEAGINOUS PLANTS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: YOUNG & THOMPSON  
; STREET: 745 South 23rd Street  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/776,210  
; FILING DATE: 24-JAN-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 94 09272  
; FILING DATE: 25-JUL-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/FR95/00957  
; FILING DATE: 18-JUL-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PATCH, Andrew J  
; REGISTRATION NUMBER: 32,925



ADDRESSEE: CUSHMAN DABY & CUSHMAN, L.L.P.  
 STREET: 1100 New York Avenue, N.W.  
 CITY: Washington  
 STATE: D. C.  
 COUNTRY: U.S.A.  
 ZIP: 20005-3918  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/362,525  
 FILING DATE: 04-JAN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: EP 92202080.5  
 FILING DATE: 08-JUL-1992  
 APPLICATION DATA:  
 APPLICATION NUMBER: EP 92203899.7  
 FILING DATE: 14-DEC-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/EP93/01763  
 FILING DATE: 07-JUL-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: KOKULIS, PAUL N.  
 REGISTRATION NUMBER: 16,773  
 REFERENCE/DOCKET NUMBER: 213289/T7020 (V)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 861-3000  
 TELEFAX: (202) 822-0944  
 TELEX: 6714627 CUSH  
 INFORMATION FOR SEQ ID NO: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1828 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA to mRNA.  
 ORIGINAL SOURCE:  
 ORGANISM: Geotrichum candidum  
 STRAIN: CMICC 335426  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 40..1731  
 OTHER INFORMATION: /product= "lipase"  
 FEATURE:  
 NAME/KEY: sig\_peptide  
 LOCATION: 40..96  
 FEATURE:  
 NAME/KEY: mat\_peptide  
 LOCATION: 97..1728  
 OTHER INFORMATION: /product= "lipase"  
 OTHER INFORMATION: /gene= "lipB"  
 US-08-362-525-11

Alignment Scores:  
 Pred. No.: 1,4e-114 Length: 1828  
 Score: 1107.00 Matches: 240  
 Percent Similarity: 56.60% Conservatives: 73  
 Best Local Similarity: 43.40% Mismatches: 196  
 Query Match: 38.65% Indels: 44  
 DB: 3 Gaps: 9

US-09-943-857-4 (1-547) x US-08-362-525-11 (1-1828)

QY 49 LeuArgPheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSer 68  
 |||||  
 DB 205 TTGGGGTTCAAGCACCCCGCCCTTTCTAGTGAATCTTACAGAGGTCTTAAAGCCCAACGAC 264  
 |||||  
 QY 69 TyrGlyProSerCysMetGlnGlnAsnProGluGlyThrPheGlu- 83  
 |||||  
 DB 265 TTCAGCTCTGCTTGTATGCAGCTTGATGCTCTGGCAATGCCCTTTCTTTCTTTGTCGTGACAAAGTA 324  
 |||||  
 QY 84 -----GluAsnLeuGlyLysThrAlaLeuAspLeuValMet 95  
 |||||  
 DB 325 GTGGGCTTGGGAAAGATTCTCTCGATATACTTAGAGGCCCTCTTTATGACATGCC- 381  
 |||||  
 QY 96 GlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnVal 115  
 |||||  
 DB 382 -----CAGGGTAGTGTCTCCATGATGAGGACTGTCTTACTTAAAGCTT 426  
 |||||  
 QY 116 ValArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGly 135  
 |||||  
 DB 427 TTCGGCCCGCTGGCACCAAGCTGATGCTAAGCTCCCGCTCATGGTTTGGATTACGGT 486  
 |||||  
 QY 136 GlyGlyPheGluLeuGlySerProThrIlePheProAlaGlnMetValThrLysSer 155  
 |||||  
 DB 487 GGTGCCCTTTGTGTGGTTCTTCTGCTTCTTACCGTGTAAACGGCTACGTCAGGAGAGT 546  
 |||||  
 QY 156 ValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrpGly 175  
 |||||  
 DB 547 GTGGAATGGCCAGCTGTGTGTGTGTTCATCACTACCTACCGCTACGGCCCTATGGA 606  
 |||||  
 QY 176 PheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGln 195  
 |||||  
 DB 607 TTCTTGGGTGGTGTATGCATCACCCTGAGGCAACACCAACGCTGTCTGCACGACGAG 666  
 |||||  
 QY 196 ArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLys 215  
 |||||  
 DB 667 CGCAAGGGTCTCGAGTGGGTAGCGACAAATTCGCCAATTTGTGTGTGATCCCGACAAG 726  
 |||||  
 QY 216 ValThrIlePheGlyGluSerAlaGlySerMetSerValLeuLeuCysHisLeuIleTrpAsn 235  
 |||||  
 DB 727 GTCATGATTTTCGTGAGTCCGCTGGTGGTGGTGTCTGCTCACCAGCTTGTGTGCTAC 786  
 |||||  
 QY 236 AspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGly 255  
 |||||  
 DB 787 GGTGGTGACAAACACCTACAAACGAAAGAGAGCTTTTCCACCTCTGCCATTCTTCACTGGC 846  
 |||||  
 QY 256 AlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluLeuLeuThrAspPhe 275  
 |||||  
 DB 847 GGTCTCTTCTTCTTACTTCTGCTTCTGTTGGTCCGAGAGTGTGCTACAGCAGATTT 906  
 |||||  
 QY 276 ValSerSerAlaGlyCys-----GlySerAlaSerAspLysLeuAlaCysLeuArg 292  
 |||||  
 DB 907 GCTCAGTATGCCGATGTACACAGCTGCGAGTGAATGACACTCTGGCTTGTCTCCGC 966  
 |||||  
 QY 293 SerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn----- 305  
 |||||  
 DB 967 AGCAAGTCCAGCGATGCTTTCACAGTGCAGAGTCTCGTATGATCTTAAAGACCTGTTT 1026  
 |||||  
 QY 306 -----ThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgPro 323  
 |||||  
 DB 1027 GGTCTGCTCCCTCAATTCTCTTGGATTGTT-----CCAGACCC 1065  
 |||||  
 QY 324 AspGlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSer 343  
 |||||  
 DB 1066 GACGGCAACATTTATCCCGCTTATGAGCTCTACCGAGCGGTAGATACGCCAAG 1125  
 |||||  
 QY 344 ValProValIleIleGlyAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeu 363  
 |||||  
 DB 1126 GTTCCCTACATTACTGGCAACCGAGGAGTATGAGGTACTATTCTTCCCGCCCTTGTATT 1185  
 |||||  
 QY 364 AsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHisAlaSer 383  
 |||||  
 DB 1186 AATGTACCACTACTCCCCCATGTTAAGAAGTGGTGAAGTACATTTGTAGCCAGGCTTCT 1245

```

QY 384 AspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerPro 403
Db 1246 GACGCTTCGCTGTGATCGTGTTCCTGCTACCCCGCTCTTGGTGGAGGTTTACCA 1305
QY 404 PheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeu 423
Db 1306 TTCGCACTGGTATCTTAATGCTCTTACCCCTCAGTTCAGCGCATGCTGCCATTTTC 1365
QY 424 GlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHisPheGlnGlyThr 443
Db 1366 ACTGATTTGCTTTTCCAGTCTCTCGTGTGTTATGCTTAACGCTACCAAGCAGCTCAAC 1425
QY 444 LysTyrSerPheLeuSerLysGlnLeuSerGlyLeu---ProIleMetGlyThrPheHis 462
Db 1426 CGCTGGACCTTACCTTGCACCCAGCTCCATCACTCTGTTCCATTTTGGGTACTTCCAT 1485
QY 463 AlaAsnAspIleValTrpGlnAspTyrLeu---LeuGlySerGlySerValIleTyrAsn 481
Db 1486 GCGAGTGATCTCTTTTCAATACTACTACGTGACCTTGGCCCATCTTCTGCT---TACCGC 1542
QY 482 AsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeuLeuValAsn 501
Db 1543 CGTACTTTTATCTCGTTTGGCAACACCACGACCCCAACGTTGGTACCACTCCCAACAG 1602
QY 502 TrpProLysTyrThrSerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAlaLeu 521
Db 1603 TGGGATATGTAACAT-----CATGAGCAAGGAGATGCTTCAGATTTCATATGATT 1653
QY 522 GlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGly 534
Db 1654 GGTAACCTATGAGAACTGACGACTTTAGAAATCGAGGGA 1692

```

# RESULT 3

```

US-08-776-210-5
; Sequence 5, Application US/08776210
; Patent No. 5942659
; GENERAL INFORMATION:
; APPLICANT: ALIBERT, Gilbert
; APPLICANT: MOULOUNGUI, Zephirin
; APPLICANT: BOUDET, Alain
; TITLE OF INVENTION: PROCESS FOR PRODUCING FATTY ACIDS OR
; DERIVATIVES THEREOF FROM OLEAGINOUS PLANTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSER: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/776,210
; FILING DATE: 24-JAN-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 94 09272
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA: WO PCT/FR95/00957
; APPLICATION NUMBER: 18-JUL-1995
; FILING DATE: 18-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: IN 387 - BE 6996
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON

```

```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1635 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-776-210-5

Alignment Scores:
Pred. No.: 1,25e-112 Length: 1635
Score: 1089.00 Matches: 238
Percent Similarity: 56.86% Conservative: 73
Best Local Similarity: 43.51% Mismatches: 192
Query Match: 38.02% Indels: 44
DB: 2 Gaps: 10

US-09-943-857-4 (1-547) x US-08-776-210-5 (1-1635)
QY 15 ProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleAsnGlu 34
Db 7 CCCACGGCGCTTCTTAATGGCAACGAGGTCACTCTGCTGTCTTGGGGGCAAGTTGAT 66
QY 35 AlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuAsnLeuArgPheLysAspPro 54
Db 67 ACCTTTAAGGGAATTCCTATTTGCTGACCTCTCTGTTGGTGTGCTTGGGTTCAAGCACCCC 126
QY 55 ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMet 74
Db 127 CAGCTTTTCACTGGATCCTACCGAGGTCTTAAGGCCAACGACTTCAGCTCTGCTGTATG 186
QY 75 GlnGlnAsnProGluGlyThrPheGlu----- 83
Db 187 CAGCTTGATCTCTGGCAATGCCATTTCTTGGCTTGCAAAAGTCGTGGGCTTGGGAAAGATT 246
QY 84 -----GluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101
Db 247 CTTCTGTATAACCTTAGAGCGCCCTCTTTATGACATGGCC-----CAG 288
QY 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgProGlyThr 121
Db 289 GGTAGTGTCTCCATGAATGAGGACGTCTCTACCTTAACGTTTCCGCCCTCTGTCACCC 348
QY 122 LysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIleGly 141
Db 349 AAGCCTGATGCTAAGCTCCCGCTCATGTTTACGTTTACGTTGGTGGTCTTGTGTGTGT 408
QY 142 SerProThrIlePheProProAlaGlnMetValThrLysSerValLeuMetGlyLysHis 161
Db 409 TCTTCTGTCTTACCCTGGTAACGGCTACGTCAAGGAGAGTGTGGAATGGGCCAGCCT 468
QY 162 IleIleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAspAsp 181
Db 469 GTTGTGTTTGTTCATCACTACCTACCGGCCCTATGATTCCTGGTGTGTGATGCC 528
QY 182 IleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrp 201
Db 529 ATCACCGCTGAGGTAACCAACCAACGCTGTCTGCACGACGACGAGGCTCTCGAGTGG 588
QY 202 ValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePheGlyGlu 221
Db 589 GTTAGCGCAACATGCCAATTTGGTGTGTGATCCCGACACAGGTTCATGATTTTCGTTGAG 648
QY 222 SerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsnThrTyr 241
Db 649 TCCGCTGTGTGCTATGATGTTGTCTACCGACTTGTTCCTACCGTGTGTGTGTGTGTGT 708
QY 242 LysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSerAsp 261
Db 709 AACGAAAGAAAGCTTTTCCACTCTGCATCTTCAGTCTGCGCGGTCTCTTCTTACTTT 768
QY 262 ProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGlyCys 281

```

```

Db 769 GACTCTACTTCTGTGTGTCGAGAGTGCCTACAGCAGATTGTCTCAGTAGTCGGGATGT 828
Qy 282 ---GlySerAlaSerAspLys-----LeuAlaCysLeuArgSerAlaSerSerAspThr 298
Db 829 GATGCCAGCCAGTCAGTGAACACTCTGGCTGTCTCCGACAGATGTCAGGATGTC 888
Qy 299 LeuLeuAspAlaThrAsnAsn-----ThrProGlyPhe 309
Db 889 TTGCACAGTCCCGAGAACTCTGACGATCTCAAGGACCTGTTTGGCCTGCTCCCTCAATTTC 948
Qy 310 LeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThr 329
Db 949 CTTGGATTGGT-----CCGAGACCCGACGCAACATTATTATCC 987
Qy 330 AspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGly 349
Db 988 GATGCGCTTATGAGCTCTACCGCAGCGGTAGATACGCCAAGGTTCCCTACATTACTGGT 1047
Qy 350 AspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAla 369
Db 1048 AACCCAGGAGATGAGGCTACTATTCTTGCCTCCCGCTTCTGCTATTATGCTACCGACTCC 1107
Qy 370 GlnAlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThr 389
Db 1108 CATGTTAAGAGTGTGAAGTACATTGTAGCGAGGCTCTGACGCTTCGCTTGATCGT 1167
Qy 390 LeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeu 409
Db 1168 GTTTTGTGCTCTACCCCGGCTCTTGTGTCGAGGGTGGCCATTCCGACTGGTATCTT 1227
Qy 410 AsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIle 429
Db 1228 AATGCTCTGACCCCTCAGTTCAGAGCGCATGCTGCCATTTCCTGATTTGCTGTCCAG 1287
Qy 430 HisAlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSer 449
Db 1288 TCTCTCGTGTGTTTGTGTTAAGCTTACCGTACCAAGAGCGTCAACCGCTGACTTACCTGGC 1347
Qy 450 LysGlnLeuSerGlyLeu---ProIleMetGlyThrPheHisAlaAsnAspIleValTrp 468
Db 1348 ACCAGCTCATAACCTCGTTTCATTTTTGGGTACTTTCCATGTGTAGTACTCTCTTTTC 1407
Qy 469 GlnAspTyrLeu---LeuGlySerGlySerValIleTyrAsnAsnAlaPheIleAlaPhe 487
Db 1408 CAATACTAGTGCACCTTGCCCATCTCTGCT---TACCGCGCTACTTTATCTCGTTT 1464
Qy 488 AlaThrAspLeuAspProAsnThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSer 507
Db 1465 GCCAACCCACGACCCCAACGTTGGCACCACCACTGAAACAGTGGGATATGTACACT--- 1521
Qy 508 SerSerGlnSerGlyAsnAsnLeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLys 527
Db 1522 -----GATGAGGCAAGGAGATGCTTCAGATTTCATATGTTGTGTAACTCTATGAGAACT 1575
Qy 528 AspAsnPheArgThrAlaGly 534
Db 1576 GACGACTTTAGATCGAGGGA 1596

RESULT 4
US-08-379-926A-7
; Sequence 7, Application US/08379926A
; Patent No. 5783414
; GENERAL INFORMATION:
; APPLICANT: CARREZ, DIRK
; APPLICANT: ROOS, JOEL
; TITLE OF INVENTION: EXPRESSION SYSTEM, INTEGRATION
; TITLE OF INVENTION: VECTOR
; TITLE OF INVENTION: AND CELL TRANSFORMED BY THIS INTEGRATION VECTOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &
; ADDRESSEE: NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400

```

```

; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/379,926A
; FILING DATE: 27-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: BE 09400102
; FILING DATE: 28-JAN-1994
; APPLICATION NUMBER: BE 09400586
; FILING DATE: 17-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: BE 09500014
; FILING DATE: 09-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 3987-13-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2045 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-379-926A-7

Alignment Scores:
Pred. No.: 1,69e-95 Length: 2045
Score: 938.50 Matches: 207
Percent Similarity: 60.81% Conservative: 80
Best Local Similarity: 43.86% Mismatches: 162
Query Match: 32.77% Indels: 24
DB: 1 Gaps: 8

US-09-943-857-4 (1-547) x US-08-379-926A-7 (1-2045)
Qy 89 ThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGlu 108
Db 13 TCGGTCAATGGGCTCTCGCTGATCTCCCTTTGGTACAGTCGGCTACCAATGCTCTCGAG 72
Qy 109 AspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeuPro 128
Db 73 GATTGCTGAACATTGACATTCCGGCTCGCGCGGGACCAACCGGAGCTCGAAGCTGCCT 132
Qy 129 ValMetLeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePhePro 148
Db 133 GTGCTGGTCTGGATCTTTGGCGAGGCTTTGAACCTTGGTTCAAGGCGATGATGATGCT 192
Qy 149 AlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaValAsn 168
Db 193 ACAACGATGATATCATCGTCGATACAGAACATCGCTATCGTGTGTTGTAGCAATGAAT 252
Qy 169 TyrArgValAlaSerTrpGlyPheLeuAlaGlyAspIleLysAlaGluGlySerGly 188
Db 253 TATCGGTGGGAGGTTTCGGGTCTTGTCCCGAAAGGAGATCTCGGAGACGGTCCGCG 312
Qy 189 AsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGly 208
Db 313 AACCTAGGCGCTCTCGGACCAACGCTCTGCCCTGTCAGTGGGTTGCCGCAACATCGAGGCC 372

```

209 PheGlyCysAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerVal 228  
Db TTTGGTGGACCCGGACAGGTGACGATTTGGGAGAAATCAGCAGGAGCCATTGCTTG 432  
229 -LeuGlyHisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheAr 248  
Db ACTAGATGACTTGTAC-----GACGGAACATCATTACAGGATAAGCCCTTGTTCG 486  
248 gAlaGlyLeuMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGl 268  
Db GGGGGCCCATGACTCGGTAGTGTGTTCCCGCAGACCCCTCGATGGGGTCAAGGG 546  
268 yAsnGluIleTrpAspLeuPheValSerAlaGlyCysGlySerAlaSerAspLysLe 288  
Db ACAGCAAGTATATGCGGTAGTGGATCTGCGAGGCTGTTCTTCTTAACGACACCT 606  
288 uAlaCysLeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGl 308  
Db AGCTTGTCTGGTGAACCTAGACTACCGGACTTCTCAATCGGCAAACTCGTGCCAGG 666  
308 yPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIl 328  
Db CATTTTAAGCTACCATTCCTGGCGTTATCATATGTGCTCGACCGGACGGCGCTT 726  
328 eThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIl 348  
Db GTGGCATCACCGGACGTTTGGGCAAGCAGGGAATATGCTCGGTCCGTTTCATCGT 786  
348 eGlyAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrAs 368  
Db GGGCGCACCAAGAGGATGAGGAGCCTTATTCGCTTGTTCAGTCCAACTTACGACGT 846  
368 nAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHis---AlaSerAspAlaGluIl 387  
Db CGACGAGGTGTCGACTACCTGGCCTCATCTTCTTATGACCTAGCCGAGACGACT 906  
387 eAspThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerProPhe-AspThrG 407  
Db TGAAGAACTAGTGGCCCTGTACCCAGACACACGACGACGAGTCTCGTTTCAGACAGG 966  
407 lYValLeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuA 427  
Db CGGCC--AACAACTGGTATCGCAATTAAGCGATTTGGCCGCAATCTTCGGCGACTTG 1023  
427 lApheIleHisAlaArgArgTyrPheLeuAsnHisPheGlnGlyThr-----AlaG 443  
1024 TCTTTCAC-CATTACCGCGGGCATTCCTCTGTATGAGAGGAATCTCCCTGATCTTC 1082  
444 --LysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThrPheHisA 463  
Db CGAACTGGTCTGACTGGGACCTATGACTATGGCACCCCACTTCTGGGGACCTTCCACG 1142  
463 laAsnAspIleValTrpGlnAspTyr-----LeuLeuGlySerGlySerV 478  
Db GAAGTGACCTGCTGACAGGTGTTCTATGGGATCAAGCCAACTATGAGTAGTCTTAGCC 1202  
478 alIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThr---AlaG 497  
Db TATCTGAGCTTTGTATACGTGGATCGATCCGAATCTCCAACTCCACCGGG 1253  
497 lYLeuLeuValAsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMetM 517  
Db GGGAGTACATTGAGTGGCGCAGTGAAGGAATCGGGCAG-----TTGATGA 1301  
517 etIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAspA 537  
Db ATTTCGGAGCGAAGCAGCGCAGTCTCTTACCGATGATTTCCGCAACGGGACATATGAGT 1361  
537 lAlaLeuMetThrAsnProSerSerPhePheVal 547  
Db TCATCTCGAGAATACCGCGCGTTCACATC 1393

RESULT 5

US-07-846-181-6  
Sequence 6, Application US/07846181  
Patent No. 5360732  
GENERAL INFORMATION:  
APPLICANT: BERKA, RANDY M  
APPLICANT: FOWLER, TIMOTHY  
APPLICANT: REY, MICHAEL W  
TITLE OF INVENTION: PRODUCTION OF ASPERGILLUS NIGER  
TITLE OF INVENTION: CATALASE-R  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENECOR INTERNATIONAL, INC.  
STREET: 180 KIMBALL WAY  
CITY: SOUTH SAN FRANCISCO  
STATE: CA  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/846.181  
FILING DATE: 19920304  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: HORN MB, MARGARET A  
REGISTRATION NUMBER: 33401  
REFERENCE/DOCKET NUMBER: GC204-US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-742-7536  
TELEFAX: 415-742-7217  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8533 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-07-846-181-6  
Alignment Scores:  
Pred. No.: 6.63e-60 Length: 8533  
Score: 632.00 Matches: 198  
Percent Similarity: 59.12% Conservative: 71  
Best Local Similarity: 43.53% Mismatches: 157  
Query Match: 22.07% Indels: 41  
DB: 1 Gaps: 10  
US-09-943-857-4 (1-547) x US-07-846-181-6 (1-8533)  
QY 108 GluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeu 127  
Db 12 GAGGATTGCTGAAATTCACATTCGGCGCCCGAGCAACCCCACTCGCGCGAATGC 71  
QY 128 ProValMetLeuTrpIlePheGlyGlyPheGlyGluIleGlySerProThrIlePhePro 147  
Db 72 CCGTGTGTCTCGATCTTTCGGCGAGGCTTTGAACCTTTCGAAAGCCCATGTATGAC 131  
QY 148 ProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaVal 167  
Db 132 GGCACACACATGGTATCATCTGCTCGATAGACAGAAATATCCCTATCGTGTGTAGCGATG 191  
QY 168 AsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySer 187  
Db 192 AACTATCGGTAGGGGCTTCGGGTTTCGCCGGAAGAGAGATTCTGGAGGACGGGTCC 251  
QY 188 GlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAla 207  
Db 252 GCCAACTTAGTCTTT--GACCAGGCTTTG---CCCTAGTGGGTGCCGCAACATCGAG 306  
QY 208 GlyPheGlyGlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSer 227



```

Db 307 GCGTTTGGTGAGA-CGAGACAAGGTGACATCTGGGAGAAATCAGACGGGCTATTCT 365
Qy 228 ValLeuCyHisLeuIleTyrAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPhe 247
Db 366 GTCTTGAT-CAGATGATCG-TAGACGGAACATCGCTTACAGGACAGCCCTTGTT 423
Qy 248 ArgAlaGlyIleMetGlnSerGlyAlaMet-ValProSerAspProValAspGlyThrTy 267
Db 424 CGGGA-GCCATCATGAGCTCGGT---ATGTGTTCCCGCAGACCTGTGCGACGGGTCAA 479
Qy 267 xGlyAsn-GluIleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspL 287
Db 480 GGGATCAGCAGATATGATGCGGTGTGGACTCTGCAGGCTGTCTCTTCCACGACA 539
Qy 287 ySLeuAlaCySLeuSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrp 307
Db 540 CCCTGGCTGTCTCGGTGAGTACACTACCGGATCTCATGCGGAACTCTGGT-C 598
Qy 307 roGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysA 327
Db 599 CGGGGATCTAGGTATC--ACCGTGGCTATCATATGCTCGCTCGACGACGGAC-G 655
Qy 327 snIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValI 347
Db 656 GCATTTCGCGGTGCGCAGATTTCGGTAAAGCAGGGAAGTATGCGGGTCCCATCA 715
Qy 347 leIleGlyAspGlnAsnAspGlyThrIlePheGlyLeuSerSerLeuAsnValThrt 367
Db 716 TCGTGGCGACCAAGAGATGAGGGACCTTGTCGCTGTT-TCAGTC---CTTACGA 771
Qy 367 hrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHis---AlaSerAspAlaG 386
Db 772 CGATCGACGAGGTAGTCGACTATCTGGGCACCTACTTCTTCTATGACGCTAGCCAGGAC 831
Qy 386 luleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspT 406
Db 832 AGCTTGAGAAATAGTGGCCCTGTATCCACGACACCAACATATGATGGGTCTCCCTTCAGGA 891
Qy 406 hrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspL 426
Db 892 CGGGCAG-GCCACAACACTGTATCCGCAATTAAGCATGGCGGCCATCTCGGGACT 950
Qy 426 euAlaPheIleHisAlaArgTyrPheLeuAsnHisPheGlnGlyThr-----443
Db 951 TGGTCTTCACATACCCGCGGCA-TTCCTGTATATGACAGGAGCTCTCCCTCGACC 1009
Qy 444 -----LysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThrPheH 462
Db 1010 TCCGGAATGGTGTACCTCGGACCTATGACTATGGCAG-CCAATTCTGGGACCTTCC 1068
Qy 462 isAlaAsnAspIleValTyrPheGlnAspTyr-----LeuLeuGlySerGlyS 477
Db 1069 ATGGAAGTACCTGCTGCAGGTCTTATGGGATCAAGCGAACTATGACGCGGATTCGA 1128
Qy 477 erValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThr---A 496
Db 1129 GCACACGAT-----TATCTGATTTGTATACAGCTGGATCCGAATCCCAATC 1179
Qy 496 laGlyLeuLeuValAsnThrProLysTyrThrSerSer---SerGln-SerGlyAsnAsn 514
Db 1180 GGGGGAGGTACATGGAAATGCCCGACGTGGCAGCCGACAGTTGATGAATTCGAGCGAAC 1239
Qy 515 LeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGly 534
Db 1240 AGCAGTCTCTTACG-----GATATTTCGCAACGGGACA 1275
Qy 535 TyrAspAlaLeuMetThrAsnProSerSerPhePheVal 547
Db 1276 TATGAGTTTATCTCGCAGAATACCGCGGCTTCCACATC 1314

```

RESULT 6

US-07-845-989-6

```

; Sequence 6, Application US/07845989
; Patent No. 5360901
; GENERAL INFORMATION:
; APPLICANT: BERKA, RANDY M
; APPLICANT: FOWLER, TIMOTHY
; APPLICANT: REY, MICHAEL W
; TITLE OF INVENTION: PRODUCTION OF ASPERGILLUS NIGER
; TITLE OF INVENTION: CATALASE-R
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENENCOR INTERNATIONAL, INC.
; STREET: 190 KIMBALL WAY
; CITY: SOUTH SAN FRANCISCO
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/845,989
; FILING DATE: 19920304
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HORN Ms, MARGARET A
; REGISTRATION NUMBER: 33401
; REFERENCE/DOCKET NUMBER: GC208-US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-742-7536
; TELEFAX: 415-742-7217
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8533 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-07-845-989-6

Alignment Scores:
Pred. No.: 6.63e-60 Length: 8533
Score: 632.00 Matches: 198
Percent Similarity: 59.12% Conservative: 71
Best Local Similarity: 43.52% Mismatches: 157
Query Match: 22.07% Indels: 41
DB: 1 Gaps: 10

US-09-943-857-4 (1-547) x US-07-845-989-6 (1-8533)

Qy 108 GluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAlaAsnLeu 127
Db 12 GAGGATTTGTGAACATTTGACATTCGGCGCCAGCGAAACCCCAATCGCGCGAATGC 71
Qy 128 ProValMetLeuTrpIlePheGlyGlyPheGluIleGlySerProThrIlePhePro 147
Db 72 CCGTGTGCTCTCGGATCTTTGGCGAGGCTTTGAACCTTGGTTCAAAGGCCATGTATGAC 131
Qy 148 ProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHisValAlaVal 167
Db 132 GCACAAACGATGTTATCATCGTAGACAGAATATCCCTATCGTGTGTGTAGCGATG 191
Qy 168 AsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAspIleLysAlaGluGlySer 187
Db 192 AACTATCGGTAGGGGCTTCGGGTTCTTCGCCGGAAGAGAGATTCTGGAGGACGGGTCC 251
Qy 188 GlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAla 207
Db 252 GCCAACTTAGGTCTTT--GACCAAGCCTTG---CCCTAGTGGGTGCCCAACATCGAG 306
Qy 208 GlyPheGlyGlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSer 227

```



Db 239 ACGCACAGGCGCTCGCCGACCGGACTGATGACCGGATCTGTGCGCGCTAGC--- 295  
 Qy 79 GluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLys 98  
 Db 296 GAGGGAACC-----AACAGCAGCAGGCGCTGCGGCCCTTAGCAAC 337  
 Qy 99 ValPheGlnAlaValLeuProGln-----SerGluAspCysLeuThrIleAsnVal 115  
 Db 338 AGCTCGAGCAGCGG---CCGACAGAAACAGCGCTGCGAGGATGCTCTTCCTCAATGTC 394  
 Qy 116 ValArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGly 135  
 Db 395 GTTGCCCGCGCGCTCGTGGCAGGCGACATCTTCCGCTCGTCTACATTCACGGA 454  
 Qy 136 GlyGlyPheGluIleGlySerProThrIlePheProAlaGlnMetValThrLysSer 155  
 Db 455 GGTGGCTACGCTTCGCGGATGCGACCGCGGACGACATTGCGCGCTTCACCAAGCAC 514  
 Qy 156 ValLeuMetGlyLysHisIleHisValAlaValAsnTyrArgValAlaSerTrpGly 175  
 Db 515 ACG-----GGAACCAAGATGGTGTGTAATCTCCAGTACCGTCTCGGCGAGCTTTGGT 568  
 Qy 176 PheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGln 195  
 Db 569 TTCTCTGCTGCGCAGCAGCAGTACGAGTACGCTGTAACGACCGCGCTGCTTGACACG 628  
 Qy 196 ArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLys 215  
 Db 629 CAATTGCGCTTCAATGGGTTCAACAGCAGCTCTCGAAGTTCGGCGGCAACCCCGATCAC 688  
 Qy 216 ValThrIlePheGlyGluSerAlaGlySerMetSerValLeuLysHisLeuIleTrpAsn 235  
 Db 689 GTTACGATTTGGCGGAGTCTCGAGCGCGAGGTCCTGTTATGAACCAAGATCAATGCGAAC 748  
 Qy 236 AspGlyAspAsnThrTyrLysGly-----LysProLeuPheArgAlaGlyIle 251  
 Db 749 ---GGCGGCAACCGCTCAAGGCTCTCGTCTCAAGAGCCCTCTTCCAGCTGCGCATC 805  
 Qy 252 MetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIle 271  
 Db 806 GGCTCTCTCTCTCTCTCCCTCCACCAAGCAGTACAACTCCCTTCGCGGAGTGTCTC 865  
 Qy 272 TyrAspLeuPheValSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeu 291  
 Db 866 TACTCCCAACTCTCTCGCGGACAACTGCACCAAGCCGCTCGTCTCGCTGCTGCTC 925  
 Qy 292 ArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAla 311  
 Db 926 GAAGCTGTCGACGCTCGCGGCTC-----GCTCGCGCGGCGGTGAAGAACTCGCGCGGC 979  
 Qy 312 TyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAspAsp 331  
 Db 980 TTCCCGTTCGGGTTTGGTGTATGTCGCGTCTCGTACGGGACCTTCCTGACTGAGCGC 1039  
 Qy 332 MetTyrLysLeuValArgAspGly---LysTyrAlaSerValProValIleLeuGlyAsp 350  
 Db 1040 GGTGCTGCTCTCTCCCAAGGCGAAGAACTCAATGGCAACCTCTTACCGCGATC 1099  
 Qy 351 GlnAsn---AspGluGlyThrIlePheGlyLeuSerSerLeu---AsnValThrThrAsn 368  
 Db 1100 ACAACCTCGACGAGGATTCATTTACTGAGCCACTATTTCAGAACACGACGATCAGC 1159  
 Qy 369 AlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHisAlaSerAspAlaGluIleAsp 388  
 Db 1160 GACCACTGCGAGCGCTC-----TCCAGTTTCGAC 1189  
 Qy 389 ThrLeuMetAlaAla---TyrPro-----GlnAspIleThrGln 400  
 Db 1190 CGCTCTCTCGCGGCTCTTCCCTATCATCATCTCGGAGCGCGCGCGCGCGAAG 1249  
 Qy 401 GlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArgIleSer 420  
 Db 1250 CAGTACCCGATCTCCGACGCGCGCTCAAGGGCAACACC-----TTCTCTCGCATCTCG 1303

Qy 421 AlaValLeuGlyAspLeuAlaPheIleHisAlaArgTyrPheLeuAsnHisPheGln 440  
 Db 1304 GCGTATCGCGGACTCGACTTCGTC---TGCCCGACCTACTGG----- 1345  
 Qy 441 GlyGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeu---ProIleMetGly 459  
 Db 1346 -----ACCGCGAGGCTTCGGCTCGTCCGCCCAAGGCGCTCTTCGACTACGCGCGC 1399  
 Qy 460 ThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySer----- 475  
 Db 1400 GCTCACACCGACCGAC-----AACTCGTACTACATCGGCTCCATCTGGAACGGC 1450  
 Qy 476 -----GlySerVal---IleTyrAsnAsnAlaPheIleAlaPheAlaThrAsp 490  
 Db 1451 AAGAGTCTGCTCTCGTCCGTCAGTCTTCAGCGCGCGCTCGCGGCTTCATCGAGACG 1510  
 Qy 491 LeuAspPro-----AsnThrAlaGlyLeuLeuValAsn-----TrpProLysTyrThr 506  
 Db 1511 TTCAACCGCAACAACACGCTGCCAACAAGACCATCAACCTTACTGCGCGACGTTGAC 1570  
 Qy 507 SerSerSerGln 510  
 Db 1571 TCGGCGCAAGCAG 1582

## RESULT 8

US-07-732-962A-1  
 ; Sequence 1, Application US/07732962A  
 ; Patent No. 5248604  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fischer, Meir  
 ; TITLE OF INVENTION: EXPRESSION OF ENZYMATICALLY ACTIVE  
 ; TITLE OF INVENTION: RECOMBINANT HUMAN ACETYLCHOLINESTERASE  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: John P. White, Esq.  
 ; STREET: 30 Rockefeller Plaza  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10112  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/732,962A  
 ; FILING DATE: 19910722  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 39304/JPW/LSW  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 977-9550  
 ; TELEFAX: (212) 664-0525  
 ; TELEX: 422523 COOP UI  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1845 base pairs  
 ; TYPE: NUCLEIC ACID  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..1842  
 ; US-07-732-962A-1

Alignment Scores: 1.02e-40 Length: 1845  
 Pred. No.: 452.00 Matches: 156

Percent Similarity: 44.59%      Conservative: 83  
 Best Local Similarity: 29.10%      Mismatches: 179  
 Query Match: 15.78%      Indels: 118  
 DB: 1      Gaps: 26

US-09-943-857-4 (1-547) x US-07-732-962A-1 (1-1845)

```

Qy 35 AlaPheLeuGlyLeuProPheAlaGluProValGlyAsnLeuArgPheLeuAspPro 54
Db 184 GCATTCTGTCGTCATCCCTTTGCGAGCCACCATGGAGCCCGCTGCTTCTGCCACCG 243
Qy 55 -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
Db 244 GAGCCCAAGACGCTTGTGTCAGGGGTGTAGACGCTACACCTTCCAGAGTGC----- 297
Qy 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluAsn 85
Db 298 ---TGCTACCAATATGTGGACACCTATACCAGGTTTGTAGGGGACCCAGATGTGGAAC 354
Qy 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
Db 355 CCCAACCGTGAGCTG----- 369
Qy 106 GlnSerGluAspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAla 125
Db 370 ---AGCGAGGACTGCTGCTACTCTAAC---GTGTGGACACCATACCCCGGCTTACATCC 423
Qy 126 AsnLeuProValMetLeuThrIlePheGlyGlyPheGluIleGlySerProThrIle 145
Db 424 CCCACCCCTGCTCTGCTGATCTATGGGGTGGCTTCTACAGTGGGCTCTCTCTCTG 483
Qy 146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHis--- 164
Db 484 -----GACGTGTACGATGGCCGCTTCTGTGTACAGGCC 516
Qy 165 -----ValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179
Db 517 GAGAGACTGCTGCTGTGTCATGAACTACCGGTGGGAGCCCTTGGCTTCTGCGCCCTG 576
Qy 180 AspAspIleLysAlaGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199
Db 577 CCGGGGAGCCGA---GAGGCCCGGCAATGTGGGTCTCTCGATCAGAGGCTGCCCCCTG 633
Qy 200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePhe 219
Db 634 CAGTGGGTGACAGAGAACGTGGAGCTTCGGGGGTGACCCACATCATGTCAGCTGTT 693
Qy 220 GlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239
Db 694 GGGGAGAGCGGGAGCCGCTCGTGGGCATGCACCTGCTG----- 735
Qy 240 ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
Db 736 TCCCCCGCCAGCGGGGCTGTTTCCACAGGGCGCTGCTGACAGCGGTGCC-----CCC 789
Qy 260 SerAspProValAspGlyThrTyrGly-----AsnGluIle 271
Db 790 AATGGACCTGG---GCCAGGTGGGCATGGGAGGCCGCTGCGAGGCCGACGAGCTG 846
Qy 272 TyrAspLeuPheValSerSerAlaGlyCys-----GlySerAlaSer 285
Db 847 GCCACCTT-----GTGGGTGTCCTCCAGCGGCACTGCTGGGAATGACACA 894
Qy 286 AspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn 305
Db 895 GAGCTGTAGCTGCTGCTTCCGACACGACGACGACGACGCTGCTGCTG-----AACCAC 945
Qy 306 ThrProGlyPheLeuAlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAsp 324
Db 946 GAATGGCAGCTGCTGCTCAGAAAGCGTCTTCGGTTCCTCTGCTGCTGCTGCTGCTGCT 1005
Qy 325 GlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerVal 344

```

```

Db 1006 GGAGACTTCTCAGTGACACCCAGAGGCCCTCATCAACGGGGAGACTTCCACGGCCTG 1065
Qy 345 ProValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe----- 358
Db 1066 CAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1125
Qy 359 GlyLeuSerSerLeuAsn-----ValThrAsnAlaGlnAlaArgAlaTyrPheLys 376
Db 1126 GGCTTCAGCAAGACAAACAGAGTCTCTCATCAGCGCGGCGGAGTTCCTGCGCGGGTGG 1185
Qy 377 GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr----- 394
Db 1186 GTCGGGGTCCCCAGGTAAGTACCTGCGAGCCGAGGCTGCTGCTGCTGCTGCTGCTGCTG 1245
Qy 395 -----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal 408
Db 1246 TGGCTGATCCCGAGGACCCCGCAGCTGAGGGAGGCCCTGAGCGATGTGTGGCGGAC 1305
Qy 409 LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe 428
Db 1306 CACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1353
Qy 429 IleHisAlaArg-----ArgTyrPheLeuAsnHisPheGlnGly----- 441
Db 1354 CAGGTGCGCGGGTCTACGCTAGCTTGTGAACACCGTGTCTCCAGCTCTCTCTGCGCC 1413
Qy 442 -----GlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIle--- 457
Db 1414 CTGTGGATGGGGTGGCGGCTACGAGATCGAGTTCATCTTTGGGATGCTGCTGCTGCTG 1473
Qy 458 ---MetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGly 476
Db 1474 CCCTCTCGAACTACACGCGCAGAGGAGAAATCTTTCGCCAGCGACTGATG----- 1524
Qy 477 SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496
Db 1525 -----CGTACTGGCCAACTTGGCCCGCAGAGGGGATCCCAATGAGCC 1569
Qy 497 -----GlyLeuValAsnTrpProLysTyrThrSerSerGln 510
Db 1570 CGAGACCCCAAGGCCCAATGGCCCGTACACGGGGGGCTCAG 1617

```

# RESULT 9

PCT-US92-06106-1  
 ; Sequence 1, Application PC/TUS9206106  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fischer, Melir  
 ; TITLE OF INVENTION: ENZYMATICALLY ACTIVE RECOMBINANT HUMAN  
 ; TITLE OF INVENTION: ACETYLCHOLINESTERASE AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: John P. White, Esq.  
 ; STREET: 30 Rockefeller Plaza  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10112  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US92/06106  
 ; FILING DATE: 19920722  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 39304-B-PCT/JPW/EAB  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 977-9550  
 ; TELEFAX: (212) 664-0525



Fri Aug 6 10:53:11 2004

CITY: Farmington Hills  
STATE: Michigan  
COUNTRY: US  
ZIP: 48334  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/318,826A  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohn, Kenneth I.  
REGISTRATION NUMBER: 30,955  
REFERENCE/DOCKET NUMBER: 2391.00001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (248) 539-5050  
TELEFAX: (248) 539-5055  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2256 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens

FEATURE:  
OTHER INFORMATION: /note= "Splice variant: Exons 1, 2,  
OTHER INFORMATION: 3, 4 and 6"  
US-08-318-826A-5

Alignment Scores:  
Pred. No.: 1,438-40 Length: 2256  
Score: 452.00 Matches: 156  
Percent Similarity: 44.53% Conservative: 83  
Best Local Similarity: 29.10% Mismatches: 179  
Query Match: 15.78% Indels: 118  
Gaps: 26

US-09-943-857-4 (1-547) x US-08-318-826A-5 (1-2256)

QY 35 AlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro 54  
DB 343 GTTTCCTGGGATCCCTTTGGGAGCCACCCATGGGACCCCGCTCGCTTTCTGCCACCG 402  
QY 55 -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71  
DB 403 GAGCCCAACGACGCTTGGTCAGGGGTGGTAGAGCTCAACCTTCAGAGTGTCTC----- 456  
QY 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85  
DB 457 ----TGCTACCAATATGTGACACCCCTATACCCAGGTTTGGGGCCACCGAGATGTGAAC 513  
QY 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105  
DB 514 CCCAACCGTGAGCTG-----GAGCTGTACGATGGCCGCTTCTTTGGTACAGGCC 675

QY 165 -----ValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179  
DB 676 GAGAGGACTGTGCTGGTGTCCATGAACTACCGGGTGGAGCCTTTGCTTCTCGCCCTG 735  
QY 180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199  
DB 736 CCGGGGAGCCGA---GAGGCCCGGGCAATGTGGGTCTCTCTGGATCAGAGGCTGCCCTG 792  
QY 200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePhe 219  
DB 793 CAGTGGGTGACGAGAACGTGGGAGCCTTCGGGGGTGACCCGACATCAGTCAAGCTGTT 852  
QY 220 GlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239  
DB 853 GGGGAGAGCGCGGAGCGCTCGTGGCATGACCTGCTG----- 894  
QY 240 ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259  
DB 895 TCCCCGCCACCGCGGCTGTTCACAGGCCGTGTGTCAGAGCGGTGCC-----CCC 948  
QY 260 SerAspProValAspGlyThrTyrGly-----AsnGluIle 271  
DB 949 AATGGACCTGG---GCCACGCTGGCATGGAGAGGCCCGTCGCGAGGCCACGACGCTG 1005  
QY 272 TyrAspLeuPheValSerSerAlaGlyCys-----GlySerAlaSer 285  
DB 1006 GCCCACCTT-----GTGGCTGTCTCTCCAGCGGACGCTGGTGGGAATACACA 1053  
QY 286 AspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn 305  
DB 1054 GAGCTGTAGCTGCTTCGGACACGACCCAGCGGAGGTCTCTGGTG-----AACCA 1104  
QY 306 ThrProGlyPheLeuAlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAsp 324  
DB 1105 GAATGGCAGCTGTCTCTCAAGAAAGCTTCTCCGGTTCTCTCTCGTCTGTGTAGAT 1164  
QY 325 GlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerVal 344  
DB 1165 GGAGACTTCTCAGTGACACCCAGAGGCCCTCATCAACGGGGAGACTTCCACGGCCCTG 1224  
QY 345 ProValIleIleGlyAspGlnAsnAspGlyThrIlePhe----- 358  
DB 1225 CAGGTGCTGGTGGTGTGAAGATGAGGCTCGTATTCTGTGTTTACGGGGCCCA 1284  
QY 359 GlyLeuSerSerLeuAsn-----ValThrAsnAlaGlnAlaArgAlaTyrPheLys 376  
DB 1285 GGCTTCAGCAAGACACGAGTCTCTCATCAGCGGGCCGAGTCTCTGCGGGGTGCGG 1344  
QY 377 GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr----- 394  
DB 1345 GTCGGGTTTCCCGAGTAACTGACCTGGCAGCGGAGGCTGTGCTCTGCTATACAGAC 1404  
QY 395 -----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal 408  
DB 1405 TGGCTGCATCCCGAGACCCCGCAGCCCTGAGGGAGCCCTGAGCGATGTGGGGCGAC 1464  
QY 409 LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe 428  
DB 1465 CACATGTGCTGTGCCCC-----GTGGCCAGCTGGCTGGGCGACTGGTGC 1512  
QY 429 IleHisAlaArg-----ArgTyrPheLeuAsnHisPheGlnGly----- 441  
DB 1513 CAGGGTCCCGGCTCTACGCTTGTAAACACCGTGTCTCCAGCTCTCTCTGGCCC 1572  
QY 442 -----GlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIle--- 457  
DB 1573 CTGTGATGGGGTGGCCACCGCTACGAGTCAAGTTCATCTTTGGATCCCTCGGAC 1632  
QY 458 ---MetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGly 476  
DB 1633 CCCTCTCGAAACTACACGGCAGAGGAGAAATCTCGCCACGAGCTGATG----- 1683

QY 477 SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496  
 Db 1684 -----CGATACTGGGCCCAACTTTGGCCGCACAGGGGATCCCAATGAGCCC 1728  
 QY 497 -----GlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510  
 Db 1729 CGAGACCCCAAGGCCCAACATGGCCCGTACAGCGCGGGGCTCAG 1776

## RESULT 11

US-08-370-156-1  
 ; Sequence 1, Application US/08370156  
 ; Patent No. 5932780  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Soreq, Hermona  
 ; APPLICANT: Zakut, Haim  
 ; APPLICANT: Shani, Moshe  
 ; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR  
 ; TITLE OF INVENTION: ANTICHOLINESTERASE SUBSTANCES  
 ; NUMBER OF SEQUENCES: 27  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Reising, Ethington, Barnard & Perry  
 ; STREET: P.O. Box 4390  
 ; CITY: Troy  
 ; STATE: Michigan  
 ; COUNTRY: US  
 ; ZIP: 48099  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/370,156  
 ; FILING DATE:  
 ; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kohn, Kenneth I.  
 ; REGISTRATION NUMBER: 30,955  
 ; REFERENCE/DOCKET NUMBER: P-307 (Mulford)  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (810) 689-3500  
 ; TELEFAX: (810) 689-4071  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2256 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-370-156-1

Alignment Scores:  
 Pred. No.: 1,43e-40 Length: 2256  
 Score: 452.00 Matches: 156  
 Percent Similarity: 44.59% Conservative: 83  
 Best Local Similarity: 29.10% Mismatches: 179  
 Query Match: 15.78% Indels: 118  
 DB: 2 Gaps: 26

US-09-943-857-4 (1-547) x US-08-370-156-1 (1-2256)

QY 35 AlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspPro 54  
 Db 343 GCTTTCTGGGATCCCTTTGGGAGCCACCCATGGGAGCCCGCTTCTGCCACCG 402  
 QY 55 -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71  
 Db 403 GAGCCCAAGCAGCCTTGGTCAGGGGTGGTAGCGCTACACCTTCAGAGTGTC----- 456  
 QY 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85  
 Db 457 ---TGCTACCAATATGTGGACACCCCTATACCCAGGTTTGGAGGGACCGAGATGTGMAAC 513  
 QY 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105

Db 514 CCCAACCGTGAGCTG----- 528  
 QY 106 GlnSerGluAspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAla 125  
 Db 529 ---AGCAGGAGTCCCTGTACCTCAAC---GTGTGGACACCATACCCCGCCCTACATCC 582  
 QY 126 AsnLeuProValMetLeuTyrIlePheGlyGlyGlyPheGluIleGlySerProThrIle 145  
 Db 583 CCACCCCTGCTCTGCTGTGATCTATGGGGTGGCTTCTACAGTGGGGCTCTCTCCCTG 642  
 QY 146 PheProProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHis--- 164  
 Db 643 -----GACGTGACGATGCCCGCTTCTGGTACAGGCC 675  
 QY 165 -----ValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179  
 Db 676 GAGAGGACTGTGCTGTGCTCACTACCGGTGGAGCCTTGGCTTCTCTGGCCCTG 735  
 QY 180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199  
 Db 736 CCGGGAGCCGA---GAGGCCCGGCAATGTGGTCTCTCTGATCAGAGCTGGCCCTG 792  
 QY 200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePhe 219  
 Db 793 CAGTGGGTGCAGGAGAACGTCGTCGCGGGTGACCCGACATCAGTACGCTGTTT 852  
 QY 220 GlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239  
 Db 853 GGGGAGAGCGCGGAGCCGCTCGTGGGCATGCACCTGCTG----- 894  
 QY 240 ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259  
 Db 895 TCCCGCCCGAGCCGGGCTGTTCACAGGGCGCTGCTGCAGAGCGGTGCC-----CCC 948  
 QY 260 SerAspProValAspGlyThrTyrGly-----AsnGluIle 271  
 Db 949 AATGAGCCCTGG---GCCAGGTGGGATGGAGAGGCCCGTCGAGGGCCAGCAGCTG 1005  
 QY 272 TyrAspLeuPheValSerSerAlaGlyCys-----GlySerAlaSer 285  
 Db 1006 GCCCACCCTT-----GTGGGCTGTCTCCAGCGGCCTGTGTGGGAATGACACA 1053  
 QY 286 AspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn 305  
 Db 1054 GAGCTGTAGCTTCCTTCGACACACCGACCGAGGCTCTGTGTG-----AACAC 1104  
 QY 306 ThrProGlyPheLeuAlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAsp 324  
 Db 1105 GAATGGCAGCTGCTGCCTCAAGAAAGGCTTCTCCGTTCTCTCTGCTGCTGTGGTAGAT 1164  
 QY 325 GlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerVal 344  
 Db 1165 GGAGACTTCTCAGTGACACCCCGAGAGCCCTCATCAACGCGGAGACTTCCACGCGCTG 1224  
 QY 345 ProValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe----- 358  
 Db 1225 CAGTGTCTGGTGGTGTGGATGAGGCTCGTATTTTCTGTTTACGGGGGCCCA 1284  
 QY 359 GlyLeuSerSerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLys 376  
 Db 1285 GGCTTCAGCAAGACACACGAGTCTCTCATCGCCGGCGCCGAGTCTCGCCGGGGTGGCG 1344  
 QY 377 GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaIleTyr----- 394  
 Db 1345 GTCGGGTTTCCCGAGTAACTGACCTGGCAGCGAGGCTGTGTCTGTGCATATACAGAC 1404  
 QY 395 -----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal 408  
 Db 1405 TGGCTGATCCCGAGACCCCGGACCGCTGAGGGAGGCCCTGAGCGATGTGGTGGCGAC 1464  
 QY 409 LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe 428

Db 1465 CACAAATGCTGTGCCCC-----GTGGCCCGAGCTGGCTGGGCGAGCTGGCTGCC 1512  
 Qy 429 ILeHisAlaArg-----ArgTyrPheLeuAsnHisPheGlnGly----- 441  
 Db 1513 CAGGGTGCCTGGCTACGCTAGCTCTTTTGAACACCGCTGCTCCACGCTCTCTCTGGGCC 1572  
 Qy 442 -----GlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProile--- 457  
 Db 1573 CTGTGGATGGGGGCCCGCTACGAGATCGAGTTCATCTTTGGATCCCTCGGAC 1632  
 Qy 458 ---MetGlyThrPheHisAlaAsnAspIleValTyrGlnAspTyrLeuLeuGlySerGly 476  
 Db 1633 CCCTCTCGAACTACACGCGAGAGAGAAATCTTCGCCCGCAGCGACTGATG----- 1683  
 Qy 477 SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496  
 Db 1684 -----CGATACTGGGCCCACTTTGCCCGCACAGGGGATCCCAATGAGGCC 1728  
 Qy 497 -----GlyLeuLeuValAsnTyrProLysTyrThrSerSerSerGln 510  
 Db 1729 CGAGACCCCAAGGCCCAATGGCCCCCGTACACGGGGGGCTCAG 1776

RESULT 12

US-08-814-095-1  
 ; Sequence 1, Application US/08814095  
 ; Patent No. 6025183  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Soreq, Hermona  
 ; APPLICANT: Zakut, Haim  
 ; APPLICANT: Shani, Moshe  
 ; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR  
 ; TITLE OF INVENTION: ANTI-CHOLINESTERASE SUBSTANCES  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: KOHN & ASSOCIATES  
 ; STREET: 30500 No. 6025183thwestern Highway, Suite 410  
 ; CITY: Farmington Hills  
 ; STATE: Michigan  
 ; COUNTRY: U.S.  
 ; ZIP: 48334  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/814,095  
 ; FILING DATE:  
 ; CLASSIFICATION: 800  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Montgomery, Ilene N.  
 ; REGISTRATION NUMBER: 38,972  
 ; REFERENCE/DOCKET NUMBER: 2391.00066  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (248) 539-5050  
 ; TELEFAX: (248) 539-5055  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2256 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: other nucleic acid  
 ; DESCRIPTION: /desc = "ACHE gene comprising exons  
 ; DESCRIPTION: 2, 3, 4 and 6"  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Homo sapiens  
 ; US-08-814-095-1

Alignment Scores: 1.43e-40 Length: 2256  
 Pred. No.:

Score: 452.00 Matches: 156  
 Percent Similarity: 44.59% Conservative: 83  
 Best Local Similarity: 29.10% Mismatches: 179  
 Query Match: 15.78% Indels: 118  
 DB: 3 Gaps: 26  
 US-09-943-857-4 (1-547) x US-08-814-095-1 (1-2256)  
 Qy 35 AlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspPro 54  
 Db 343 GCTTTCTGGCATCCCTTTTCGGAGCCACCCATGGAGCCCGCTTCGCTTCGTCACCG 402  
 Qy 55 -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71  
 Db 403 GAGCCCAAGCAGCCTTGTGTAGGGTGTATCACTCAACCTTCCAGAGTGC----- 456  
 Qy 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85  
 Db 457 ---TGCTACCAATATGTGGACACCTATACCCAGGTTTGTAGGCGCACCGATGTGGAAC 513  
 Qy 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105  
 Db 514 CCCAACCGTGAGCTG----- 528  
 Qy 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125  
 Db 529 ---AGCGAGGACTGCTGTACTCAAC---GTGTGGACACCATACCCCGGCGCTACATCC 582  
 Qy 126 AsnLeuProValMetLeuTyrIlePheGlyGlyPheGluIleGlySerProThrIle 145  
 Db 583 CCCACCCCTGCTGCTGTGATCTATGGGGTGGCTTCTACAGTGGGGCGCTCTCCCTTG 642  
 Qy 146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHis--- 164  
 Db 643 -----GAGGTGTACGATGGCGCTTCTGTGTACAGGCC 675  
 Qy 165 -----ValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179  
 Db 676 GAGGAGACTGTGCTGTCCATGAACACCGGGTGGGAGCCTTTGGCTTCTGGCCCTG 735  
 Qy 180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199  
 Db 736 CCGGGAGCCGA---GAGCCCGGGCAATGTGGGTCTCTCTGATCAGAGGCTGGCCCTG 792  
 Qy 200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe 219  
 Db 793 CAGTGGTGCAGGAGACGTGGCAGCTTCGGGGGTGACCCGACATCAGTACGCTGTTT 852  
 Qy 220 GlyLysSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239  
 Db 853 GGGGAGAGCGCGGAGCCGCTCGTGGGCATGCACCTGCTG----- 894  
 Qy 240 ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259  
 Db 895 TCCCGCCCGCAGCGGGGCTGTTTCCACAGGCGCGTGTCTCAGAGCGGTGCC----- 948  
 Qy 260 SerAspProValAspGlyThrTyrGly-----AsnGluIle 271  
 Db 949 AATGGACCTGG---GCCACGGTGGGCATGGAGAGGCCCGTGGAGGCCACGCGAGCTG 1005  
 Qy 272 TyrAspLeuPheValSerSerAlaGlyCys-----GlySerAlaSer 285  
 Db 1006 GCCCACCTT-----GTGGGCTGTCTCCAGCGGCGACTGTGGGGAATGACACA 1053  
 Qy 286 AspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn 305  
 Db 1054 GAGCTGTAGCTGCTTTCGGACACCGACCGCAGCGCAGTCTCTGTG-----AACAC 1104  
 Qy 306 ThrProGlyPheLeuAlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAsp 324  
 Db 1105 GAATGCGACGTGCTGCTCAAGAAAGCGTCTTCCGGTTCCTCTGCTGTGGGTAGAT 1164  
 Qy 325 GlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerVal 344





```
QY 382 AlaSerAspAlaGluIleAspThrLeuMetAlaIleTyr----- 394
Db 1240 GTGAGTGGATTGGAAAGAAATCCATCTTTTTCATTACAGACTGGGTAGATGATCAG 1299
QY 395 ---ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThr 413
Db 1300 AGACTGAAACTACCGTGAGCCCTTGGGTGATGTTGTTGGGATTATAATTCATATGC 1359
QY 414 Pro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAla 431
Db 1360 CTGCTGCTGGAGTTCCACCAAGAAGTTCTCAGAAATGGGGAAATAATGCCTTTTTC 1413
QY 432 ArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGln 451
Db 1414 -----TACTATTTTGAACAC-----CGATCTCCCAA 1440
QY 452 LeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyr 471
Db 1441 CTTCGTCGCCAGAATGATGGAGTGATGATGATGATGATGATGATGATGATGATGATGAT 1491
QY 472 LeuLeuGly-----Ser 475
Db 1492 GTCTTTGTTTACCTCTCGAAAGAGAGATAATTACACAAAGCCGAGAAATTTTGTAGT 1551
QY 476 GlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsn--- 494
Db 1552 AGATCCATAGT-----AAACGGTGGGCAAAATTTTGCAAAATATGGGAATCCAAATGAG 1605
QY 495 ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db 1606 ACTCAGAACATAGCACAGCTGGCTGTCTTCAAAAGCACTGAACAA 1653

RESULT 14
US-08-318-826A-7
; Sequence 7, Application US/08318826A
; Patent No. 5891725
; GENERAL INFORMATION:
; APPLICANT: Soreq, Hermona
; APPLICANT: Zakut, Haim
; APPLICANT: Eckstein, Fritz
; TITLE OF INVENTION: Synthetic Antisense
; TITLE OF INVENTION: Oligodeoxynucleotides and Pharmaceutical Compositions
; TITLE OF INVENTION: Containing Them
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 5891725thwestern Hwy., Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: US
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/318,826A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: 2391.00001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (248) 539-5050
; TELEFAX: (248) 539-5055
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3016 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 160..2010
; OTHER INFORMATION: /note= "Splice Variant: Exons 1, 2,
; OTHER INFORMATION: 3, 4, 5 and 6"
US-08-318-826A-7

Alignment Scores: 2,32e-40 Length: 3016
Pred. No.: 452.00 Matches: 156
Score: 44.59% Conservative: 83
Percent Similarity: 29.10% Mismatches: 179
Best Local Similarity: 15.78% Indels: 118
Query Match: 2 Gaps: 26
DB:

US-09-943-857-4 (1-547) x US-08-318-826A-7 (1-3016)
QY 35 AlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspPro 54
Db 343 GCTTTCTGGCATCCCTTTTGGGAGCCACCCATGGGAGCCCGCTGCTTTCTGCCACCG 402
QY 55 -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
Db 403 GAGCCCAAGCAGCCTTGTGTAGGGTGTAGACGCTCACACCTTCACAGATGTC----- 456
QY 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85
Db 457 ---TGCTACCAATATGTGGACACCCCTATACCCAGGCTTTTGAGGGCAGCAGATGTGGAC 513
QY 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
Db 514 CCCAACCGTGAGCTG----- 528
QY 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125
Db 529 ---ACGAGGAGTGTCTGTACTCAAC---GTGTGGACACCATACCCCGGCTCATATCC 582
QY 126 AsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIleGlySerProThrIle 145
Db 583 CCCACCCCTGCTCTGATCTATGGGTGCTTCTACAGTGGGGCTCTCCCTCTTG 642
QY 146 PheProProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHis--- 164
Db 643 -----GACGTGTAGTGGCCGCTTCTGTGTACAGGCC 675
QY 165 -----ValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179
Db 676 GAGAGGAGTGTCTGTGTCCATGAATACCGGTGGGAGCCTTTGGCTTCTGTGGCCCTG 735
QY 180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199
Db 736 CCGGGGAGCCGA---GAGGCCCGGGCAATGGGTCTCTCGATCAGAGGCTGGCCCTG 792
QY 200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe 219
Db 793 CAGTGGGTGCAGGAGAACCGTGGCAGCCTTCCGGGGGTGACCCACATCATGTAGCTGTT 852
QY 220 GlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239
Db 853 GGGGAGAGCGGGAGCCCTCGTGGGCATGTCACCTGCTG----- 894
QY 240 ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
Db 895 TCCCCGCCCGCGGGGCTGTTCACAGGGCGGTGCTGTCAGAGCGGTGCC-----CCC 948
QY 260 SerAspProValAspGlyThrTyrGly-----AsnGluIle 271
Db 948
```

```
Db 949 AATGACACCTGG---GCCACGGTGGGATGGAGAGGCCCGTCGACGGCCACCGCAGCTG 1005
Qy 272 TyrAspLeuPheValSerSerAlaGlyCys-----GlySerAlaSer 285
Db 1006 GCCACACCTT-----GTGGCTGTCTCCAGCGCGCACCTGGTGGGAATGACACA 1053
Qy 286 AspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn 305
Db 1054 GAGTGTGTAGCTGCTTTCGACACGACGACGAGCGAGTCTCTGGTG-----AACCAC 1104
Qy 306 ThrProGlyPheLeuAlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAsp 324
Db 1105 GAATGGCAGCTGCTCCCTCAAGAAAGCTCTCCGCTTCTCTTCGTCCTGGTGTAGAT 1164
Qy 325 GlyLysAsnLeuThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerVal 344
Db 1165 GGAGACTTCTTCAGTGACACCCACGAGGCCCTCATCAACGGGGAGACTTCCACGGCCTG 1224
Qy 345 ProValIleIleGlyAspGlnAsuAspGluGlyThrIlePhe----- 358
Db 1225 CAGGTGCTGTGGTGTGGTGAAGGATGAGGCTCGTATTTCTGTTTACGGGGCCCA 1284
Qy 359 GlyLeuSerSerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLys 376
Db 1285 GGCTTCAGCAAGACAAGCTCTCTCATCGCGGGCGAGTTCCTGGCGGGGTGCGG 1344
Qy 377 GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr----- 394
Db 1345 GTCGGGTTCCTCCAGGTAAAGTGACCTGCGACCGCGCTGAGGAGGCCCTGAGCGATGTGGTGGCGAC 1404
Qy 395 -----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal 408
Db 1405 TGGCTGCATCCGAGGACCGCGCACGCTGAGGAGGCCCTGAGCGATGTGGTGGCGAC 1464
Qy 409 LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe 428
Db 1465 CACATGTGTGTGCCCC-----GTGCGCCAGCTGGCTGGCGGACTGGCTGCC 1512
Qy 429 IleHisAlaArg-----ArgTyrPheLeuAsnHisPheGlnGly----- 441
Db 1513 CAGGGTCCCGGGTCTACGGCTAGCTTTTGACACCGCTGCTCCACGCTCTCTCTGGCCC 1572
Qy 442 -----GlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIle--- 457
Db 1573 CTGTGGATGGGGTGGCCCGACGGCTACGAGATCGAGTTCATCTTTGGGATCCCGCTGGAC 1632
Qy 458 ---MetGlyThrPheHisAlaAsnAspIleValTyrPheGlnAspTyrLeuLeuGlySerGly 476
Db 1633 CCCTCTCGAACTACACGGCAGAGGAGAAATCTTCGCCCGACGAGCTGATG----- 1683
Qy 477 SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496
Db 1684 -----CGATACTGGGCCAACTTTGGCCCGCACAGGGGATCCCAATGAGCCC 1728
Qy 497 -----GlyLeuValAsnTyrProLysTyrThrSerSerSerGln 510
Db 1729 CGAGACCCCAAGGCCCCCAATAGCCCGCCCTACACGGCGGGGCTCAG 1776

RESULT 15
US-08-370-156-5
; Sequence 5, Application US/08370156
; Patent No. 5932780
; GENERAL INFORMATION:
; APPLICANT: Soreq, Hermona
; APPLICANT: Zakut, Haim
; APPLICANT: Shani, Moshe
; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
; TITLE OF INVENTION: ANTICHOLINESTERASE SUBSTANCES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Reising, Ethington, Barnard & Perry
; STREET: P.O. Box 4390
; CITY: Troy
```

```
; STATE: Michigan
; COUNTRY: US
; ZIP: 48099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,156
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: P-307 (Mulford)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 689-3500
; TELEFAX: (810) 689-4071
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3016 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: 160..2010
; US-08-370-156-5
```

```
Alignment Scores:
Pred. No.: 2,32e-40 Length: 3016
Score: 452.00 Matches: 156
Percent Similarity: 44.59% Conservative: 83
Best Local Similarity: 29.10% Mismatches: 179
Query Match: 15.78% Indels: 118
DB: 2 Gaps: 26
```

```
US-09-943-857-4 (1-547) x US-08-370-156-5 (1-3016)
```

```
Qy 35 AlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro 54
Db 343 GCTTTCCTGGCATCCCTTTGGGAGCCACCCATGGGACCCCGTCGCTTTCTGCCACCG 402
Qy 55 -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
Db 403 GAGCCCAAGCAGCCTTGGTCAGGGGTGTAGACGCTACACCTTCCAGAGTGTC----- 456
Qy 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85
Db 457 ---TGCTACCAATATGTGGACACCCCTATACCCAGGTTTGGAGGCGACCGAGATGTGAAC 513
Qy 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
Db 514 CCCAACCGTGAGCTG----- 528
Qy 106 GlnSerGluAspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAla 125
Db 529 ---AGGAGAGACTGCTGTACCTCAAC---GTGTGGACACCATACCCCGGCTCATATCC 582
Qy 126 AsnLeuProValMetLeuTyrIlePheGlyGlyGlyPheGluIleGlySerProThrIle 145
Db 583 CCCACCCCTGCTCGTCTGGATCTATGGGGTGGCTTCTACAGTGGGGCCTCTCTCTG 642
Qy 146 PheProProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHis--- 164
Db 643 -----GACGTGTACGATGGCGCTTCTTGGTACAGGCC 675
Qy 165 -----ValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179
Db 676 GAGAGACTGTGCTGTGTCCATGAACCTACCGGGTGGGAGCCTTTGGCTTCTCGGCCCTG 735
Qy 180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199
```

Db 1729 CGAGACCCCAAGGCCCCCAATGCCCCGTACACGGGGGGGGCTCAG 1776

Search completed: August 5, 2004, 21:21:04  
Job time : 132 secs

```

736 CCGGGAGAGCCGA---GAGGCCCCGGGCAATGTGGGTCTCTGGATCAGAGGCTGCCCTG 792
QY
200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe 219
Db
793 CAGTGGGTGCAGGAGAACGTGGCAGCTTCCGGGGGTGACCCGACATCAGTGACGCTGTT 852
QY
220 GlyGluSerAlaGlySerMetSerValLeuLeuCysHisIleuIleTrpAsnAspGlyAspAsn 239
Db
853 GGGGAGAGAGCGGGAGCGCTCGTGGGCATGACCTGCTG----- 894
QY
240 ThrTyrlsGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
Db
895 TCCCGCCCGCCAGCGGGCTGTTTCCACAGGGCCGTGCTGCAGAGCGGTGCC-----CCC 948
QY
260 SerAspProValAspGlyThrTyrls-----AsnGluIle 271
Db
949 AATGGACCCCTGG---GCCACGGTGGGCATGGGAGAGGCCCGTGCAGGGGCCACGCGCTG 1005
QY
272 TyrAspLeuPheValSerSerAlaGlyCys-----GlySerAlaSer 285
Db
1006 GCCCACCTT-----GTGGGCTGTCTCCAGGGGGCAGCTGGTGGGAATGACACA 1053
QY
286 AspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn 305
Db
1054 GAGCTGGTAGCTGCTTCGACACAGCAGCGCAGGTCTCTGTG-----AACCAC 1104
QY
306 ThrProGlyPheLeuAlaTyrlsSerSerLeu---ArgLeuSerTyrlsLeuProArgProAsp 324
Db
1105 GAATGGCAGCTGTGCCTCAAGAAAGCTCTTCCGGTTCTCTCGTCTGCTGTGTAGAT 1164
QY
325 GlyLysAsnIleThrAspAspMetTyrlsLeuValArgAspGlyLysTyrlsAlaSerVal 344
Db
1165 GGAGACTTCTCAGTGACACCCCGCAGGCGCTCATCAACGGGGAGACTTCCACGGCCCTG 1224
QY
345 ProValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe----- 358
Db
1225 CAGGTGCTGGTGGGTGGTGAAGGATGAGGCTCGTATTTCTGTTTACGGGGCCCCA 1284
QY
359 GlyLeuSerSerLeuAsn-----ValThrAsnAlaGlnAlaArgAlaTyrlsPheLys 376
Db
1285 GGCCTTCAGCAAGACAAAGAGTCTCTCATCAGCGGGCGGAGTTCCTGGCCGGGGTGGG 1344
QY
377 GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrls----- 394
Db
1345 GTCGGGGTTCGCCAGTAAGTACCTGGCAGCGCGGCTGTGTCTGCTGCTATTACACAGAC 1404
QY
395 -----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal 408
Db
1405 TGGCTGCATCCCGAGGACCCCGCACCGCTGAGGGAGGCCCTGAGCGATGTGTGGCGGAC 1464
QY
409 LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe 428
Db
1465 CACAAATGCTGTGCCCC-----GTGGCCAGCTGGCTGGCGACTGGCTGCC 1512
QY
429 IleHisAlaArg-----ArgTyrlsPheLeuAsnHisPheGlnGly----- 441
Db
1513 CAGGGTGCCCGGGTCTACGCCTACGCTTTGAAACACCGTGCTTCCAGGCTCTCCTGGCCC 1572
QY
442 -----GlyThrLysTyrlsSerPheLeuSerLysGlnLeuSerGlyLeuProIle--- 457
Db
1573 CTGTGGATGGGGTGGCCCGCAGCGCTACGAGATCGAGTTCATCTTTGGGATCCCGCTGGAC 1632
QY
458 ---MetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrlsLeuLeuGlySerGly 476
Db
1633 CCCTCTCGAAACTACACGGCAGAGAGAAAATCTTCGCCAGCGACTGATG----- 1683
QY
477 SerValIleTyrlsAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496
Db
1684 -----CGTACTGGGCCAACTTTGCCCGCACAGGGGATCCCATGAGCCC 1728
QY
497 -----GlyLeuLeuValAsnTrpProLysTyrlsSerSerSerGln 510

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 5, 2004, 20:33:42 ; Search time 515 Seconds

(without alignments)

5207.801 Million cell updates/sec

Title: US-09-943-857-4

Perfect score: 2864

Sequence: 1 SMNSRCPAGRLGVSPTAKLA.....DNFRTAGYDALMTNPSPFFV 547

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09943857/runat\_29072004\_090902\_29857/app\_query.fasta\_1.711  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNIT8=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09943857 @CGN 1.1.354 @runat\_29072004\_090902\_29857  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NRG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2:	/cgn2_6/ptodata/2/pubpna/PTC_NEW_PUB.seq:
3:	/cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq:
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6:	/cgn2_6/ptodata/2/pubpna/PTCUS_PUBCOMB.seq:
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9:	/cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10:	/cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11:	/cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
14:	/cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15:	/cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16:	/cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
------------	-------	-------	--------	-------	-------------

1	2392.5	83.5	1532	10	US-09-943-857-3	Sequence 3, Appli
2	2034	71.0	1541	10	US-09-943-857-7	Sequence 7, Appli
3	2020	70.5	1548	10	US-09-943-857-5	Sequence 5, Appli
4	1903	66.4	1511	10	US-09-943-857-9	Sequence 9, Appli
5	1875.5	65.5	1469	10	US-09-943-857-1	Sequence 1, Appli
6	1156.5	40.4	1687	16	US-10-369-493-27923	Sequence 27923, A
7	456	15.9	2416	15	US-10-032-233-39	Sequence 39, Appli
8	456	15.9	2416	17	US-10-324-466-39	Sequence 39, Appli
9	456	15.9	2416	17	US-10-413-432-39	Sequence 39, Appli
10	455	15.9	2416	9	US-09-748-739A-7	Sequence 7, Appli
11	452	15.8	1725	9	US-09-810-861B-5	Sequence 5, Appli
12	452	15.8	2416	15	US-10-032-233-33	Sequence 33, Appli
13	452	15.8	2416	17	US-10-324-466-33	Sequence 33, Appli
14	452	15.8	2416	17	US-10-413-432-33	Sequence 33, Appli
15	452	15.8	5767	9	US-09-810-861B-3	Sequence 3, Appli
16	452	15.8	14446	9	US-09-810-861B-4	Sequence 4, Appli
17	451	15.7	1725	16	US-10-326-892-1	Sequence 1, Appli
18	451	15.7	2381	9	US-09-880-107-2271	Sequence 2271, Ap
19	451	15.7	2416	9	US-09-748-739A-5	Sequence 5, Appli
20	451	15.7	2416	9	US-09-748-739A-16	Sequence 16, Appli
21	451	15.7	2416	15	US-10-032-233-21	Sequence 21, Appli
22	451	15.7	2416	15	US-10-032-233-23	Sequence 23, Appli
23	451	15.7	2416	15	US-10-032-233-25	Sequence 25, Appli
24	451	15.7	2416	15	US-10-032-233-43	Sequence 43, Appli
25	451	15.7	2416	17	US-10-324-466-21	Sequence 21, Appli
26	451	15.7	2416	17	US-10-324-466-23	Sequence 23, Appli
27	451	15.7	2416	17	US-10-324-466-25	Sequence 25, Appli
28	451	15.7	2416	17	US-10-413-432-23	Sequence 23, Appli
29	451	15.7	2416	17	US-10-413-432-21	Sequence 21, Appli
30	451	15.7	2416	17	US-10-413-432-25	Sequence 25, Appli
31	451	15.7	2416	17	US-10-413-432-43	Sequence 43, Appli
32	451	15.7	2416	17	US-10-413-432-31	Sequence 31, Appli
33	451	15.7	2444	9	US-09-893-519A-112	Sequence 112, App
34	450	15.7	2416	9	US-09-748-739A-3	Sequence 3, Appli
35	450	15.7	2416	15	US-10-032-233-27	Sequence 27, Appli
36	450	15.7	2416	15	US-10-032-233-31	Sequence 31, Appli
37	450	15.7	2416	15	US-10-032-233-35	Sequence 35, Appli
38	450	15.7	2416	15	US-10-032-233-37	Sequence 37, Appli
39	450	15.7	2416	17	US-10-324-466-31	Sequence 31, Appli
40	450	15.7	2416	17	US-10-324-466-35	Sequence 35, Appli
41	450	15.7	2416	17	US-10-324-466-37	Sequence 37, Appli
42	450	15.7	2416	17	US-10-413-432-27	Sequence 27, Appli
43	450	15.7	2416	17	US-10-413-432-31	Sequence 31, Appli
44	450	15.7	2416	17	US-10-413-432-35	Sequence 35, Appli
45	450	15.7	2416	17	US-10-413-432-35	Sequence 35, Appli

#### ALIGNMENTS

#### RESULT 1

US-09-943-857-3  
; Sequence 3, Application US/09943857  
; Publication No. US20030124701A1  
; GENERAL INFORMATION:  
; APPLICANT: Shaw, Jei-Fu  
; APPLICANT: Lee, Guan-Chiun  
; APPLICANT: Tang, Shye-Jye  
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES  
; FILE REFERENCE: 08919-066001  
; CURRENT APPLICATION NUMBER: US/09/943,857  
; CURRENT FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1532  
; TYPE: DNA  
; ORGANISM: Candida rugosa  
US-09-943-857-3

Alignment Scores: 7.83e-279 Length: 1532  
Pred. No.: 2392.50 Matches: 502  
Score:

Fri Aug 6 10:53:12 2004

Percent Similarity:	94.37%	Conservative:	1
Best Local Similarity:	94.18%	Mismatches:	9
Query Match:	83.54%	Indels:	26
DB:	10	Gaps:	16
US-09-943-857-4 (1-547) x US-09-943-857-3 (1-1532)			
QY	15	ProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleLeuAsnGlu	34
DB	2	CCACCGCCCAAGCTCCCAACGCGGACACCATCACCGGTCTCAACGCCATCATCAACGAG	61
QY	35	AlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro	54
DB	62	CGGTCTCTCGCATCTCTTTCGAGCGCGCGGTGGCAACCTCCGCTTCAAGGACCT	121
QY	55	ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMet	74
DB	122	GTGCGGTACTCTGGCTCGCTCAACGGCCAGAAGTT-ACT---TACGGCCCG---TGATG	174
QY	75	GlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuVal	94
DB	175	CAGCAGAACCCCGAGGCGACGTTTGAAGAGAACCTTGGCAAGACGGCACTCGACTTGGT	234
QY	95	MetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsn	114
DB	235	ATGAGTCCCAAGGTGTTCCAGCGGTGCTTCCCGAGAGTGGAGTGCCTCACCATCAAC	294
QY	115	ValValArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuThrIlePhe	134
DB	295	GTGCTGGCGCGCGCGGACCAAGCGCGCGCCCAACCTCCCGGTCTGCTCGATCTTT	354
QY	135	GlyGlyGlyPheGluIleGlySerProThrIlePheProAlaGlnMetValThrLys	154
DB	355	GGCGGTGGGTGTGAGATCGGAGCGCCCAACCATCTTCCCTCCCGCCAGATGGTCAACAG	414
QY	155	SerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrp	174
DB	415	AGTGTGCTCATGGGCAAGCC-ATCATCCACGTGGCGGTCAACTACCGGTGTTCGTGGT	473
QY	175	GlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAsp	194
DB	474	GGGTCTTGGCTGTGTGATGACATCAAGCGCGGAGCGGAGCGCGGTGTGAAGGAC	533
QY	195	GlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSer	214
DB	534	CAGCGTTTGGGATCGAGTGGGTGGCAGACACATTCGCGGTTCGGCGGACCCGAGC	593
QY	215	LysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuLysHisIleLeuTrp	234
DB	594	AAGGTGACATC-TTTGGCGAG---GCGGCGAGCATGTCCGTGTTGCGCACCTCATCTGG	649
QY	235	AsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSer	254
DB	650	AACGACGGCGCAACACACGTACAAAGGCGAGCGGTGTTCCGCGCGGCGATCATGAG---	706
QY	255	GlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeu	274
DB	707	GGAGCCATGTGCCG---GACCCGGTGGACGACGATGCGGCAACGAGATCTACGACCTC	763
QY	275	PheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAla	294
DB	764	TTTGTCTCAGTGTGCTGTGGCGCGCGGCGAGCAAGCTCGCGTGTGGCGAGTGGC	823
QY	295	SerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSer	314
DB	824	---AGCGACACCTTGTCTGATGCCAACACACACTCTCTGGGTCTTGGCGTACTCTCG	880
QY	315	LeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLys	334
DB	881	TTCCGGTGTG-----TACTCCCGCGCGGCGGCAAGACATCACCGATGATGTACAAG	934
QY	335	LeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAspGlnAsnAspGlu	354
DB	935	TTGTGTCGCGGACGCGCAAGTATGCAAGCGTTCCTCGTATCATTTGGCGACCAAGACGAG	994
QY	355	GlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyr	374
DB	995	GGCACCATCTTTGGCTC-----TTGAACGTGACCAACGAATGCTCAGGCGCGGTCTAC	1047
QY	375	PheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr	394
DB	1048	TTCAACGAG---TTTCATCCACGCGGACGCGGAGATCGACACTTGTATGGCGGCGTAC	1104
QY	395	ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrPro	414
DB	1105	CCCAGAGACATCACCCAGGGT---CGTTTCGACAC---GGT-----TCAACGCTCACCCCG	1154
QY	415	GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyr	434
DB	1155	CAGTTCAAGAGAATC---GCGGTGCTCGGCGACCTTGCATTCATCCACGCGCGGCTAC	1211
QY	435	PheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGly	454
DB	1212	TTCTCTCAACACATTCACAGGCGCGGACCAAGTACTCTGTTCTCTC---AAGCAGCTC---GGG	1265
QY	455	LeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGly	474
DB	1266	TTGCCATCATGGCACCTTCCATGCCACAGCATTTGTGTGGCAGGACTACTTGTGGGA	1325
QY	475	SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsn	494
DB	1326	AGCGGACGCTCATCTACAAACGCGTTCATCGCTTCGCCACCGACTTGGACCCCAAC	1385
QY	495	ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsn	514
DB	1386	ACCGCGGGTGTGTGGTGAACCTGCCCAAGTACACACGAGC-----CAGGGCAACAAC	1439
QY	515	LeuMetMetIleAsnAlaLeuGlyLysTyrThrGlyLysAspAsnPheArgThrAlaGly	534
DB	1440	TTGATGATGATCAACGCGCTTGGGCTTGTACACCGGCAAGGACAACTTCGACCGCGCTGC	1499
QY	535	TyrAspAlaLeuMetThrAsnProSerSerPhePheVal	547
DB	1500	TACGACGCTTGTGATGACCAACCCG-----TTCTTTGTG	1532
RESULT 2			
US-09-943-857-7			
; Sequence 7, Application US/09943857			
; Publication No. US20030124701A1			
; GENERAL INFORMATION:			
; APPLICANT: Shaw, Jai-Fu			
; APPLICANT: Lee, Guan-Chiun			
; APPLICANT: Tang, Shye-Jye			
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES			
; FILE REFERENCE: 08919-066001			
; CURRENT APPLICATION NUMBER: US/09/943,857			
; CURRENT FILING DATE: 2001-08-31			
; NUMBER OF SEQ ID NOS: 11			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 7			
; LENGTH: 1541			
; TYPE: DNA			
; ORGANISM: Candida rugosa			
US-09-943-857-7			
Alignment Scores:			
Pred. No.:	26-235	Length:	1541
Score:	2034.00	Matches:	434
Percent Similarity:	87.57%	Conservative:	31
Best Local Similarity:	81.73%	Mismatches:	49
Query Match:	71.02%	Indels:	24
DB:	10	Gaps:	15
US-09-943-857-4 (1-547) x US-09-943-857-7 (1-1541)			
QY	17	AlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleLeuAsnGluAlaPhe	36

```

Db 7 GCACGGTCGCCAACGGCGACCATCACCGGTCTCAACGCCATCATCAACGAGCGGTC 66
Qy LeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspProValPro 56
Db CTGGCATTCCTTTGCGGAGCGCGGTGGCAACCTCCGCTTCAAGGACCTTGTGGCG 126
Qy TyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGln 76
Db TACCGTGGGTCTCTCAACGGGTCAATCTTCAACCGGTACGGTCCG---TGCAACGAG 183
Qy AsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGln 96
Db AACCCGAGGACCTACGAGGAGAACTCCCAAGGTGGCGCTTGACTTGGTATGAC 243
Qy SerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValVal 116
Db TCCAAAGGTGTTCAGGCTGTCTCCCAACAGCGAGACTGCTCACCATCAACGTTGGT 303
Qy ArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGly 136
Db CGGCGCGCGGSCCAAGCGGCGCGCAACCTCCCGGTCTGCTCTGCACTTTGGCGGT 363
Qy GlyPheGluIleGlySerProThrIlePheProGlnMetValThrLysSerVal 156
Db GGGTTTGAATCGGACGACCTACCATCTTCCCTCCCGCTCAGATGGTCTCAAGAGTGTG 423
Qy LeuMetGlyLysHisIleLeuValAlaValAsnTyrArgValAlaLaserTrpGlyPhe 176
Db CTCATGGGCGGCC-ATCATCCAGCTGGCGTCAACTACCGCTTGGCGTCTTTGGTTTC 482
Qy LeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArg 196
Db TTGGCGCGTCCGACATCAAGCGCGGCGAGCTCCAAATGCCGCGCTCAAGGACCGCG 542
Qy LeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysVal 216
Db TTGGGATGCGTGGTGGGAGACACATTTGCCGGGTTCGGCGCGCACCCGACCAAGGTG 602
Qy ThrIlePheGlyGluSerAlaGlySerMetSerValLeuLysHisLeuIleTrpAsnAsp 236
Db ACCATCTTTGGAG---GGCGGCGAGTGTCCGTGTGTGGCCACCTTCTCGAATGGC 658
Qy GlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAla 256
Db GCGCAACAACGTAACAAGCGCAAGCGGTGTTCGCGCGGCGCATCATGCG---GGAGCC 715
Qy MetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheVal 276
Db ATGGTGGCG---GACCCGGTGGACGCGACCTATGGACC-CAATCTATGACACGTTGGTG 771
Qy SerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSer 296
Db GCTACG---GGCTGCGAGAGTGCACCAACAGCTTGGCTGTGGTGGTCTT---ACT 825
Qy AspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArg 316
Db CAGGCAATTCCTCGATGCCAACCAACGACCCCTGGGTCTTGTGCGTACACCTCGTTGGCG 885
Qy LeuSerTyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLysLeuVal 336
Db TTG-----TACTCCCGGCGCGGCGCAACATCACCGGATGACATGTACAAAGTTGGTA 939
Qy ArgAspGlyLysTyrAlaSerValProValIleLeuGlyAspGlnAsnAspGluGlyThr 356
Db CGGAGCGGCAAGTATGCAAGCTTCCGTGTATCATTTGGCGACAGAACGAGGCGCTTC 999
Qy IlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLys 376
Db TTTGTTGT-CTC-----TTGAACACCAACCGGCGCGGCGGAGGATACCTCAGA 1052
Qy GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGln 396

```

```

Db 1053 AAG---TTCATCCACGCGCACCGCGGATATACCGCATTTGAAGCGCGGTACCCAGC 1109
Qy AspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPhe 416
Db GATGTACCCAGGGT---CCGTTCCGACACCGGCAATCTCAACGCGCTTACACCCAGCTC 1165
Qy LysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgAtgTyrPheLeu 436
Db AAGCGGATCAATGCTGCTTGGGACCTCACCTTTACCTCTCGCGCGCTACTTCTCTC 1226
Qy AsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuPro 456
Db AACCACTACACCGGTGGTCCCAAGTACTCGTTCTCTC---AAGCAGCTT---GGTTG-CCC 1279
Qy IleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGly 476
Db ATTCTCGGAGC-TTCCACGCGGAACGACATTTGTGTGGCAGCACTTTTGTGGGCGCGGC 1338
Qy SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496
Db AGCGTCATCTACAAACAGGTTTCATCGCTTGGCACCACCTTGGACCCCAACACCGCG 1398
Qy GlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuMet 516
Db GGC---TTGGTGCAGTGGCGCAAGTC-ACC---AGCAGCCAGCGGCGGACAACTTGTATG 1451
Qy MetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAsp 536
Db CAGTCAGTGCCTTGGGCTTGTACACCGCAAGGACAACTTCCGCGACCGCGGCTACAAC 1511
Qy AlaLeuMetThrAsnProSerSerPhePheVal 547
Db GCTTTGTTGGCGACCGG---CACTTTTTCGTG 1541

```

## RESULT 3

```

US-09-943-857-5
; Sequence 5, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jei-Fu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1548
; TYPE: DNA
; ORGANISM: Candida rugosa
US-09-943-857-5

```

## Alignment Scores:

```

Pred. No.: 1e-233          Length: 1548
Score: 2020.00           Matches: 422
Percent Similarity: 87.12%      Conservatives: 38
Best Local Similarity: 79.92%    Mismatches: 57
Query Match: 70.53%             Indels: 17
DB: 10                      Gaps: 9

```

US-09-943-857-4 (1-547) x US-09-943-857-5 (1-1548)

```

Qy 15 ProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleLeuGlu 34
Db 2 CCACTGCGCAGCTCGCAACGCGGACACCATCACCGGTCTCAACGCCATCATCAACGAG 61
Qy 35 AlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspPro 54
Db 62 GCGTTCTCGGTATTTCCCTTTGCTCAGCGCGCGGTGGGCAACCTCCGCTTCAAGCGCT 121
Qy 55 ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMet 74

```

122 GTCCGCTACGCGCTCTCAATGTCAGAGTTTACTGTATGG---CCCTG-TGCATG 177  
 75 GlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuVal 94  
 178 CAGATGAACCAATTGGGCAACTGGGACTCCTCGCTTCCCAAGGTGCGCATCAACTTG--- 234  
 95 MetGlnSerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsn 114  
 235 ATGCAGTCCCAAGCTCTTCCAGCGCGTGTCTTAAACGGCGAGGACTGTCTCACCATCAAC 294  
 115 ValValArgProProGlyThrLysAlaGlyAlaValAsnLeuProValMetLeuTrpIlePhe 134  
 295 GTGTGTCGG---CCGGGACCAACCGCGGTGTCACCTCCCGGTGATGTTGTTG 351  
 135 GlyGlyGlyPheGluIleGlySerProThrIlePheProAlaGlnMetValThrLys 154  
 352 GCGCGCGGTGTTGAGTTGGCGGTCCAGTCTCTTCCCTCCGACACAGATGATCACCGCC 411  
 155 SerValLeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrp 174  
 412 AGCGTGTCTATGGGCAACGCCCATCATCCAGTGTAGTGAATACCGGTGCTTCTGTGG 471  
 175 GlyPheLeuAlaGlyAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAsp 194  
 472 GGGTCTTGGCTGTCTCCAGATCAACGCGGAGGCGGAGCGCGGTTCGACGAC 531  
 195 GlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSer 214  
 532 CAACGTTGGTGTTCAGTGTGGTGGCGACAACTTGGCGGTTCGCGCGGACCGCTCC 591  
 215 LysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrp 234  
 592 AAGGTGACCATCTTTGGTGAGG--GCGGCGACATGTCTGTAATGTGTGTCAGTCTCTCTGG 649  
 235 AsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSer 254  
 650 AACGACGCGGCAACACGTACACGCGACGCGTGTTCGCGTCCGCGCATCATGACG-- 706  
 255 GlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeu 274  
 707 GGGGCCATGTTGCCCG--GACCGCGTGGATGGCGCTACGGCAGCAGCATCTACGACGAC 764  
 275 PheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAla 294  
 765 GTGGTGTCTACGCGCGCTGTGCGCAGTGTCCGCGCAAGCTCGCGTGTGGCA--GCA 822  
 295 SerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSer 314  
 823 TCGAAGCACAACACTCTTCCAGGCGCACCGACACTCGGGGCGCTTGGCGTACCCCTCG 882  
 315 LeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLys 334  
 883 TTGCGGTGG--TTTCTCCGCGCGCGCGCGACCTTCATCAGCATGACATGTTCAG 940  
 335 LeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAspGlnAsnAspGlu 354  
 941 TTGGTGTGCGGACGCGCAAGTGTGCAACGTTCGCGTGTATCTTGGCGACCCAGACGACGAG 1000  
 355 GlyThrIlePheGlyLysSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyr 374  
 1001 GGCACAGTGTTCGTTG-----TTGACGTGACTACCGATGCTCAGGACCGCCAGTAC 1054  
 375 PheLysGlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaIaTyr 394  
 1055 TTCAGAGAA---TTCATCCACGCGCAGCAGCGGAGATGCACACTTGTATGGCGGCGTAC 1111  
 395 ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrPro 414  
 1112 CCAGCGACATCAACCCAGGGT---CCGTTCCACCGCGCATCTTCACGCGCATCACCCCG 1168  
 415 GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyr 434

1169 CAGTTCAACCGGATTGTCAGCGGTGCTGGTGACCTTCGCTTCACTCTCCCGCGGCTAC 1228  
 435 PheLeuAsnHisPheGlnGlyGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGly 454  
 1229 TTCTCTCAACCATCTTCAGGGCGGCGCACCAAGTACTCGTTCC--TCGAAGCAGCTT---GGG 1283  
 455 LeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGly 474  
 1284 TTGCGGTGATTGGCACCCACCGCACGACATTTGTGGCAGGACTTTTGTGGTGGAGC 1343  
 475 SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsn 494  
 1344 CACAGCAGCGCGGTGTACAAACCGCTTTATGGCTTTGCCAACGACCTCGACCCGAC 1403  
 495 ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsn 514  
 1404 AAGCGCGTGTGCTTGTGTAAGTGGCCCAAGTACACGACGACG-----CAGGGCAACAAC 1457  
 515 LeuMetMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGly 534  
 1458 TTGTTGCAGATCAACGCGCTTGGGCTTGTACACCGCGCAAGGACAACTTCCCGCACCGCTGGC 1517  
 535 TyrAspAlaLeuMetThrAsnPro 542  
 1518 TACGACGCGTGTTCACACCCCG 1541  
 RESULT 4  
 US-09-943-857-9  
 ; Sequence 9, Application US/09943857  
 ; Publication No. US20030124701A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Shaw, Jai-Fu  
 ; APPLICANT: Lee, Guan-Chiun  
 ; APPLICANT: Tang, Shye-Jye  
 ; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES  
 ; FILE REFERENCE: 08919-066001  
 ; CURRENT APPLICATION NUMBER: US/09/943,857  
 ; CURRENT FILING DATE: 2001-08-31  
 ; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 9  
 ; LENGTH: 1511  
 ; TYPE: DNA  
 ; ORGANISM: Candida rugosa  
 US-09-943-857-9  
 Alignment Scores:  
 Pred. No.: 1,41e-219 Length: 1511  
 Score: 1903.00 Matches: 433  
 Percent Similarity: 86.84% Conservative: 29  
 Best Local Similarity: 81.39% Mismatches: 43  
 Query Match: 66.45% Indels: 38  
 DB: 10 Gaps: 20  
 US-09-943-857-4 (1-547) x US-09-943-857-9 (1-1511)  
 QY 17 AlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGluAlaPhe 36  
 DB 7 GCCACGCTCGCCCAACGGCGACACCATCACCGGTCTCAACGCCATCATCAACGAGCGGTTTC 66  
 QY 37 LeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspProValPro 56  
 DB 67 CTCGGCATTTCCCTTTCCGAGCGCGGTGGGCACTCCCGCTTCAGGACCCCGTGGCGG 126  
 QY 57 TyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGln 76  
 DB 127 TACTCCGCTCGCTCGATGCCAGAGTTTCAC-----TACGGCCCG---TGCATGACGAC 179  
 QY 77 AsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGln 96  
 DB 180 AACCCCGAGGCGACCTTACGAGGAGAACCTCCCGGAGGCGCTCGCTCGATTTGGTGTGAC 239  
 QY 97 SerLysValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValVal 116



```

Db      240  TCCAGAGTGTGTGAGCGGGT-----CCGAGCGAGGACTGTCTCACCATCAACGCTGGT 293
Qy      117  ArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGly 136
Db      294  CGCGCGCGCGGACCAAGCGGGTGCCAACTCCCGGTGATGCTCTGATCTTTGGCGGC 353
Qy      137  GlyPheGluLeuGlySerProThrIlePheProProAlaGlnMetValThrLysSerVal 156
Db      354  GGGTTTGGAGTGGGTGGCAGCACCTTCCTCCCGCCGAGATGATCACCAGAGCAT 413
Qy      157  LeuMetGlyLysHisIleIleHisValAlaValAsnTyrArgValAlaSerTrpGlyPhe 176
Db      414  GCCATGGCAAGCCCATCTCCAGTGGAGCGTCAACTACCGCGTGTCTGTGGGGTTC 473
Qy      177  LeuAlaGlyAspAlaGlyLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArg 196
Db      474  TTGGCTGGCGAGATCAAGCGCGAGGCGAGTGCACACGCCGGTTTGAAGGACCAAGCG 533
Qy      197  LeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysVal 216
Db      534  TG-GGCATGAGTGGGTGGCGGACCAACTTGGCGCGTTCGGCGCGACCCGACCAAGGTG 592
Qy      217  ThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAsp 236
Db      593  ACCATCTTTGGCGAG---GCGGGGAGCATGTCTGGTCAATGTGCCACATTTCTTGGAGCGAC 649
Qy      237  GlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAla 256
Db      650  GCGCAACAACGATACAGGCGAGCGCTCTTCGGCGCGGCAATCATGCAG---GGGGCC 706
Qy      257  MetValProSerAspProValAspGlyThrTyrGlyAsnGluLeuTyrAspLeuPheVal 276
Db      707  ATGGTC---CGGACGCG-GTGGACGGCTC-TACGGCAACGAGATCTTTGACCTTTGGCG 761
Qy      277  SerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSer 296
Db      762  TCAC-GCGGGCTGGCGGCGCGCGACCAAGCTTCGCGTGTGGCGGTG---AGC 817
Qy      297  AspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArg 316
Db      818  GACACGTTGGAGGACCGCCACCAACACCCCTGGGTCTTGGCGTACTCTCTGTTGGCG 877
Qy      317  Leu-SerTyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLysLeuVa 336
Db      878  TTGTACT-----CCCGCGCGCGCGCGCTGAACATCACCGACGACATGTGC---CTTGG 928
Qy      336  lArgAspGlyLysTyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyTh 356
Db      929  TCGCGGGCGAGATGTGCACT---CCTGTGATCATCGCGGACCAAGACGACGAGGGCAC 984
Qy      356  rIlePheGlyLeuSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLy 376
Db      985  CTTCTTTGCG-----ACCTTGAAGCTGACCGGAT-----GCCAGCGCCGAGTACT 1032
Qy      376  sGlnSerPheIleHisAlaSerAspAlaGluLeuAspThrLeuMetAlaTyrProGl 396
Db      1033  TCAGCAGTGTGTCCAGCGCGAGCGCGAGTGC-GACACGTTGATGACGGCGTACCC--- 1089
Qy      396  nAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPh 416
Db      1090  -GACATCACCCA-GGG---CCGTTGCACACGGGTTT-CTCAACGCGCCCTCACCGCGAGTT 1143
Qy      416  eLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgTyrPheLe 436
Db      1144  CAAGAGAATC---GCGGTCTCGGCGACCTTGTACGC-----CGTGGTACTTCTCT 1193
Qy      436  uAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuPr 456
Db      1194  CAACCACTACACCGCGCGGACCAAGTACTCAATTCCTC---AAGCAGCTC---GGCTTGC 1247
Qy      456  oIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGl 476

```

```

Db      1248  GGTGCTCGGAAGCTTCCACTCCACGACATTTGTTCCAGGACTACTTGTGGGAGCGG 1307
Qy      476  ySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAl 496
Db      1308  CTGCTCATCTACAAACACGCTTCAATGGCTTTGCCACGACTTGGACCCCAACACCGC 1367
Qy      496  aGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGlnSerGlyAsnAsnLeuWe 516
Db      1368  GGGGTGTTGGTGAAGTGGCCGAGTACACAGCAGC-----CAGGCAACAACCTTGAT 1421
Qy      516  tMetIleAsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAs 536
Db      1422  GATGATCAACGCTTGGCTTTACACCGCAAGCAACATC-CGCACCGCGCGGTACGA 1480
Qy      536  pAlaLeuMetThrAsnProSerSerPheVal 547
Db      1481  CGGTTGTTCTTCAACCCG---CCGTTCTTTGTG 1511

```

## RESULT 5

```

US-09-943-857-1
; Sequence 1, Application US/09943857
; Publication NO. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jai-Fu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1469
; TYPE: DNA
; ORGANISM: Candida rugosa
US-09-943-857-1

```

## Alignment Scores:

```

Pred. No.: 2,87e-216 Length: 1469
Score: 1875.50 Matches: 394
Percent Similarity: 80.80% Conservative: 31
Best Local Similarity: 74.90% Mismatches: 61
Query Match: 65.49% Indels: 44
DB: 10 Gaps: 16

```

US-09-943-857-4 (1-547) x US-09-943-857-1 (1-1469)

```

Qy      19  LeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGluAlaPheLeuGly 38
Db      12  CTCGCCAACGGCGACACCATCACCGTCTCAACGCCATTTGTCAACGAAAGTTTCTCGGC 71
Qy      39  IleProPheAlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSer 58
Db      72  ATACCGTTTGGCGAGCGCGCGTGGCA--CTCCGCTTCAAG-----111
Qy      59  GlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnPro 78
Db      112  -----CTCAACGGCGAGCAGTTTACC---TACGCGCCG---TGCATGCAGATGAACCT 159
Qy      79  GluGlyThrPheGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLys 98
Db      160  ATGGGCTCGTTTCA-----TTGGTCTCAGTCCAGTCCAG 191
Qy      99  ValPheGlnAlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro 118
Db      192  ATCTTCCAAGTGTGCTTCCCAACGACGAGGACTGTCTCACC-----233
Qy      119  ProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyPhe 138
Db      234  -----ACCAGGCGCAGTGGTCTCCCGGTGATGCTCTGATCTTTGGCGGTGGTTT 287
Qy      139  GluIleGlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMet 158

```

288 GAGCTTGGCGGCTCCAGCCCTTTTCAGGAGACAGATGTTGGCCAGAGCGTCTCATG 347  
 159 GlyLysHisIleIleHisValAlaValAsnTyrArgValalaSerTrpGlyPheLeuAla 178  
 348 GGTAAACCGGTGATCCACGTCGAGCATGAACTACCGCGTGGCGTCATGGGGTCTTGGCC 407  
 179 GlyAspAspIleIleHisValAlaValAsnTyrArgValalaSerTrpGlyPheLeuAla 198  
 408 GGCCCCGACATCCAGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 467  
 199 MetClnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIle 218  
 468 ATGACGATGGGTGGGGGAGCAATCTGCTGGTGGTGGGGGAGGAGGAGGAGGAGGAGGAG 527  
 219 PheGlyGluSerAlaGlySerMetSerValLeuGlyHisIleValAlaValAsnTyrArg 238  
 528 TACGGCCGAG--GGGGGAGCATGTCGACGTTTGTGACCTTGTGTGGAACGAGCGGAC 584  
 239 AsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetVal 258  
 585 AACACGTCACACGCGCAAGCGGTTGTCGCGCGCCATCATGCGAG--GGCTGCGATGGT 641  
 259 ProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSer 278  
 642 CCG--GACCCGGTGGAGCGGACGTCGACGCGGAGATCTACACCGAGTGTGGCGTCT 698  
 279 AlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAlaSerSerAspThr 298  
 699 GCCGGTGTGGCAGTGGCAGGACGACGAGCTCGCGTGTGCGCGGCTTTCTCAGGACAG 758  
 299 LeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSer 318  
 759 TTGTACACGAGGCGACGAGCAGCGCGCGGCGTGTGGCGTACCGCGTGTGGCGGTG-- 815  
 319 TyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLysLeuValArgAsp 338  
 816 TATCTCCCGCGCGCGGACGCGACCTTCATCCCGGACGATGATGCTTGGTGGCGGAC 875  
 339 GlyLysTyrAlaSerValProValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe 358  
 876 GGCAAGTACGACACGTCGCGGTGATCATCGCGGACGACGAGACGAGCGGCGCTTTGTT 935  
 359 GlyLeuSerSerLeuAsnValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSer 378  
 936 GGGCTC-----TTGAACGTGACACGATGCTCAGGCGCGGCGGCTTTCAGGAG-- 986  
 379 PheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIle 398  
 987 TTCATCCACGCGGATGCGGAGATCGACAGTTCGATGGCGGCGGTACACGAGGACATC 1046  
 399 ThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuThrProGlnPheLysArg 418  
 1047 ACCCAGGGT--CCGTTCGACACCGGATCTTCAATGCCATCACCCTCCGCGGTTCACACG 1103  
 419 IleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgArgTyrPheLeuAsnHis 438  
 1104 ATC--GGGTTCGCTTGGCGACCTTGGTTCAGCGTTCGCGTTCGCTTCTCTCACTAC 1160  
 439 PheGlnGlyGlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIleMet 458  
 1161 TACCAGGGCGGACCAAGTACTCGTCTC---AAGCAGCTT---GGGTGGCGCGGTTCACACG 1213  
 459 GlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrIleLeuGlySerGlySerVal 478  
 1214 GGCACCTTCCAGCGCAACGACATCATCTGGCAGGACTACTTGGTGGCGAGCGGCGGTG 1273  
 479 IleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAlaGlyLeu 498  
 1274 ATCTCAACACGCGTTCATTTGGCTTGGCAGGACTCTGACCCGCGACGAGGCGGCGGTG 1333  
 499 LeuValAsnTrpProLysTyrThrSerSerGlnSerGlyAsnAsnLeuMetIle 518

1334 TGGACCAACTGGCCCACTGATACACGAGCAG-----GGCAACAACCTTGATGCAGATC 1386  
 519 AsnAlaLeuGlyLeuTyrThrGlyLysAspAsnPheArgThrAlaGlyTyrAspAlaLeu 538  
 1387 AACGGCTTGGGTGTGATACCGCGGAGGACAACTTCGCCCGGATGGTACAGGCGCTC 1446  
 539 MetThrAsnProSerSer 544  
 1447 TTTTCCAAACCGCTTCT 1464  
 RESULT 6  
 US-10-369-493-27923  
 ; Sequence 27923, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 27923  
 ; LENGTH: 1687  
 ; TYPE: DNA  
 ; ORGANISM: Neurospora crassa  
 US-10-369-493-27923  
 Alignment Scores:  
 Pred. No.: 4,02e-129 Length: 1687  
 Score: 1156.50 Matches: 247  
 Percent Similarity: 60.39% Conservative: 96  
 Best Local Similarity: 43.49% Mismatches: 175  
 Query Match: 40.38% Indels: 50  
 DB: 16 Gaps: 10  
 US-09-943-857-4 (1-547) x US-10-369-493-27923 (1-1687)  
 QY 13 SerValProThrAlaLysLeuAlaAsnGlyAspThrIleThrGlyLeuAsnAlaIleIle 32  
 DB 1 GCCGCCCGGACTGTCCACCATTTCCACCGGT---ACGATTGGGTGCTTAACGGCATCCTC 57  
 QY 33 AsnGluAlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLys 52  
 DB 58 ACTGAGGCGCTTCAACGGAATCCCTTACGCGCTTCTCCGACCGGCAACCTTCGCTCAAG 117  
 QY 53 AspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyProSer 72  
 DB 118 CTTCCCGTGAGACTTAAGTCTCTCTGGGTCTTTGATGCTCTGGCATCGCCCTGCT 177  
 QY 73 CysMetGlnGlnAsnProGluGlyThrPheGluGluAsnLeuGlyLysThrAlaLeuAsp 92  
 DB 178 TSCCCCGAGTCTCTTCTGTCACACCTCGTCAACGAGTTCCTCGCTCAG---GTTATCGAT 234  
 QY 93 LeuValMetGlnSerLysValPheGlnAlaValLeuProGlnSerGlyAspCysLeuThr 112  
 DB 235 AAGATCGTTAAACACGACGCTTTTCAAGACTATCTACACTCAAGTCAAGGAGGACTGCTTAC 294  
 QY 113 IleAsnValAlaArgProGlyThrLysAlaGlyAlaAsnLeuProValMetLeuTrp 132  
 DB 295 ATCTCGCTCCTCTGTCGCGGACCGGACCGGCTGGTGAAGTCTCCCGCTCTCTTCTG 354  
 QY 133 IlePheGlyGlyPheGluIle----- 140  
 DB 355 ATCTTTGGTGGTGTTCGAAGTGAGAAATCCAGCTTATATACGCGATGTAATGAACAAG 414  
 QY 141 -----GlySerProThrIlePheProAlaGlnMetValThr 154

Db 415 TGCTAAACTTCACAGCTCGGATCGCGTCCATGTACGATGCGCTCCCTAGTCACCAA 474  
 QY 154 sSerValLeuMetGlyLysHisIleHisValAlaValAsnTyrArgValAlaSerTr 174  
 Db 475 CGCTATCAACATGGTAAAGCGGTACGTCTACGTTCGCGTCACTACCGGTGCGGTGGT 534  
 QY 174 pGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAs 194  
 Db 535 TGGTTTCATCCCGGAAAGAGATCTTAAGAGACGGCTCTTCCAACTTTGGGTACCTTGA 594  
 QY 194 pGlnArgLeuGlyMetGlnTTPValAlaAspAsnIleAlaGlyPheGlyGlyAspProSe 214  
 Db 595 CCAGCGATGGCGCTCCAGTGGTTCGCGCAACATTCGCTTCGCGGTGACCCAGA 654  
 QY 214 rLysValThrIlePheGlyGlySerAlaGlySerMetSerValLeuCysHisLeuLeTr 234  
 Db 655 CAAGGTCACTATCTGGGCGAGTCCGCGGTGCCATGTCGCTTTTCAACAGATGTCTCT 714  
 QY 234 pAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSe 254  
 Db 715 CTATGACGGTGACAAACAGCTACACGCGCAAGCCCTTTTCGTCGCGCATCATGAATC 774  
 QY 254 rGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLe 274  
 Db 775 TGGTTTCATGTCCTCCCGCGCGCTGCGTCACTGCCCAAGGCGCAGAAAGTCTACGAC 834  
 QY 274 uPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAlaCysLeuArgSerAl 294  
 Db 835 CGTGTCAAGAACCGCGTCTCTGTCGTGTGTGACACCTTCGTCGCTGCGCTCT 894  
 QY 294 aSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeuAlaTyrSerSe 314  
 Db 895 TCCTACGAGACTTTCTCAAGCGCGTAACTCCGTCCTGGGATTCGTGCTACAATC 954  
 QY 314 rLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThrAspAspMetTyrLy 334  
 Db 955 GGTGCTCTTCTTACCTCCGCGACCGCATGGCAAGCTTTGACTCAGAGCGCGGATAA 1014  
 QY 334 sLeuValArgAspGlyLysTyrAlaSerValProValIleIleGlyAspGlnAsnAspGl 354  
 Db 1015 GCTCATGCTCGTAAAGTACGCGCGGTCCCATGATCATCGCGCATCAAGAGGATGA 1074  
 QY 354 uGlyThrIlePheGlyLeuSerSerLeuAsnValThrAsnAlaGlnAlaArgAlaTy 374  
 Db 1075 GGGCACTCTTCTCCCTCTCCAGAGCAACATCACCCACCAAGCAAGCTGTGTGAGTA 1134  
 QY 374 rPheLysGlnSerPheIleHis--AlaSerAspAlaGluIleAspThrLeuMetAlaAl 1194  
 Db 1135 CCTCAAGGATATCTTCTCAAGACGCCCGAGTGCAGATTAAGTCTCTCTCTCGAC 1194  
 QY 393 aTyrProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAlaLeuTh 413  
 Db 1195 CTACAGTACCTTATCTCGCGCGTCTCGCTTTGGCAGCGGCTCTTCAACAGATTTA 1254  
 QY 413 rProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgAr 433  
 Db 1255 CCCCAGCTTCAAGCGCTGCGCGCATCTTTCGCGATCTCATCTTCAACCTCAGCGCGCG 1314  
 QY 433 gTyrPheLeuAsnHisPheGlnGlyThr-----LysTyrSe 446  
 Db 1315 CATCTTCTCGAC-----GCCCGCACCACTCTCAACCCCTCGGTGCGCGCTGGTC 1365  
 QY 446 rPheLeuSerLysGlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIl 466  
 Db 1366 GTATCTTGGTCTTACAACTTTGGCACCCATCTTGGACCTTTCAGCGCTCCGATAT 1425  
 QY 466 e-----ValTrpGlnAspTyrLeuLeuGlySerGlySerValIleTyrAsnAsnAlaPh 484  
 Db 1426 CTGCAAGGTGTTCTACGGCATCTCGCCCAACTACGCGCAGCAAAAGCATCCAGTCTTACTA 1485  
 QY 484 eIleAlaPheAlaThrAspLeuAspProAsnThrAlaGly----- 497

Db 1486 GCCCAACTTTGTTTACAACTTGACCCCAACGACGCTCCGTCGTCGCTCTAAGAG 1545  
 QY 498 ----LeuLeuValAsnTTPProLysTyrThrSerSerGlnSerGlyAsnAsnLeuMe 516  
 Db 1546 CAAGTTCAGCGAGGATTTGGCGCAATGGCAGAGGAGAGAAAG----- 1588  
 QY 516 tMetIleAsnAlaLeuGlyLeuTyrThrGly-----LysAspAsnPheArgThrAlaGl 534  
 Db 1589 -CTGGTCCAGTCTCTTTCGGACTATGCGGATATCTTCCGGATGATTTCCGCTCTGATTC 1647  
 QY 534 yTyrAspAlaLeuMetThrAsn 541  
 Db 1648 GTATAACTGGATTAAAGGCTAAT 1669  
 RESULT 7  
 US-10-032-233-39  
 ; Sequence 39, Application US/10032233  
 ; Publication No. US20030153062A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Watkins, Jeffrey D.  
 ; APPLICANT: Hancock, James D.  
 ; TITLE OF INVENTION: Butyrylcholinesterase Variants with  
 ; FILE REFERENCE: P-IX 4642  
 ; CURRENT APPLICATION NUMBER: US/10/032,233  
 ; CURRENT FILING DATE: 2001-12-20  
 ; NUMBER OF SEQ ID NOS: 50  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 39  
 ; LENGTH: 2416  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Butyrylcholinesterase variant  
 ; NAME/KEY: CDS  
 ; LOCATION: (214)...(1935)  
 US-10-032-233-39  
 Alignment Scores:  
 Pred. No.: 4,62e-44 Length: 2416  
 Score: 456.00 Matches: 160  
 Percent Similarity: 45.17% Conservative: 83  
 Best Local Similarity: 29.74% Mismatches: 185  
 Query Match: 15.92% Indels: 110  
 DB: 15 Gaps: 26  
 US-09-943-857-4 (1-547) x US-10-032-233-39 (1-2416)  
 QY 22 GlyAspThrIleThrGlyLeuAsnAlaIleAsnGluAlaPheLeuGlyIleProPhe 41  
 Db 277 GGTGGCAGGTAACA-----GCCCTTCTTGAATTCCTAT 312  
 QY 42 AlaGluProValGlyAsnLeuArgPhePheLysAspProValProTyrSerGlySerLeu 61  
 Db 313 GCACAGCCACCTCTGCTAGACTTCGATTCAAAAGCCACAGCTCTGACCAAGTGTCT 372  
 QY 62 AsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGlyThr 81  
 Db 373 GATATTGGATGTCACCAAAATATGCAAAATCTTGTGT-----CAGAACATAGATCAAAAGT 429  
 QY 82 PheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101  
 Db 430 TTTCCAGGCTTCATGGA-----TCAGAGATGTGGAAC 462  
 QY 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro---ProGly 120  
 Db 463 CCAACACACTGACCTCAGTGAAGACTGTTATATCTAATGTATGATTCACACACCTAA 522  
 QY 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIle 140  
 Db 523 CCAAAA-----AATGCCACTGATTTGATGATGATTTATGTTGTTGTTTCAAACT 573  
 QY 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160

```

Db 574 GGAACATCATCTTTA-----CATGTTTATGAGCGCAG 606
QY 161 -----HisIleHisValAlaValAsnTyrArgValAlaSerTrp 174
Db 607 TTTCTGGCTCGGTTGAAGAGTTATTGTAGTGTCAATGAAGTATAGGTGGTGGCCTA 666
QY 175 GlyPheLeuAla-----GlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeu 192
Db 667 GGAATCTTAGCTTGGCCAGGAAT-----CCTGAGGCTCCAGGGAACATGGGTTTA 717
QY 193 LysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAsp 212
Db 718 TTTGATCAACAGTTGGCTCTTCAGTGGGTTCAAAAATAATAGCAGCCTTTGGTGAAT 777
QY 213 ProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCyHisLeu 232
Db 778 CCTAAAGTGTAACTCTCTTTGGAGAAAGTGCAGGAGCAGCTTCAGTTAGCTGCAATTG 837
QY 233 IleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMet 252
Db 838 CTT---TCTCTCGGAAGCATTCA-----TTGTTCCAGAGCCATTCTG 879
QY 253 GlnSerGlyAlaMetValProSerAspProValAspGlyThrTyr-----GlyAsnGlu 270
Db 880 CAAGTGGATCATGAATGCTCTCTGGCGGTAACTCTTTATGAAGCTAGGAACAGA 939
QY 271 IleTyrAspLeuPheValSerSerAlaGlyCysGly-----SerAlaSerAspLysLeu 288
Db 940 ACGTTGAACCTA--GCTAAATTTGACTGGTCTCTAGAGAGAATGAGACTGAAATAATC 996
QY 289 AlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGly 308
Db 997 AAGTCTCTTAGAATAAAGATCCCAAGAAATCTT-----CTGAATGAAGCAATT 1047
QY 309 PheLeuAlaTyrSerSer---LeuArgLeuSerTyrLeuProArgProAspGlyLysAsn 327
Db 1048 GTTGCCCTATGGGACTCTTTGTCAGTAACTTTGGTCCGACCGTGGATGTTT 1107
QY 328 IleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValle 347
Db 1108 CTCACCTGACATGCCAGACATATTACTTGAACCTGGACAATTTAAAMAAACCCAGATTTG 1167
QY 348 IleGlyAspGlnAsnAspGluGlyThrIlePhe-----GlyLeuSer 361
Db 1168 GTGGGTGTAAATAAGATGAAGGCAGACTTTTATGCTATGCTCTCTGCTTCAGC 1227
QY 362 SerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPhe 379
Db 1228 AAGATAACAAATAGTATCACTAGAAAGAAATTTCCAGGAAGTTTAAATAATTTT 1287
QY 380 IleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr-----394
Db 1288 CCAGGAGTGAGTGAGTTGGAAGGAATCCATCTTTTTCATTACACAGACTGGGTAGAT 1347
QY 395 -----ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAla 411
Db 1348 GATCAGAGACCTGAAACTACCGTGAGGCCCTGGGTGATGTTGTTGGGATTAATATTC 1407
QY 412 LeuThrPro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIle 429
Db 1408 ATATGCTCTGCTGGAGTTCCACCAAGAGTTCTCAGAATGGGGAATAATGCTTTTTC 1467
QY 430 HisAlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSer 449
Db 1468 -----TACTATTTTGAACAC-----CGATCC 1488
QY 450 LysGlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGln 469
Db 1489 TCCAAATCTCGTGGCCAGAAATGATGGAGTGATGATGGCTATGAAAT-----1539
QY 470 AspTyrLeuLeuGly-----474

```

```

Db 1540 GAATTTGTCTTGTGTTTACCTCTGGAAAGAGAGATAATTACACAAAGCCGAGAAATT 1599
QY 475 ---SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspPro 493
Db 1600 TTGAGTAGATCCCATAGTG-----AACGGTGGGCAAAATTTTGCAGAAATATGGAATCCA 1653
QY 494 Asn---ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db 1654 AATGAGACTCAGAACAAATAGCACAAAGCTGGCTGTCTTCAAAAGCACTGAACAA 1707

RESULT 8
US-10-324-466-39
; Sequence 39, Application US/10324466
; Publication No. US20040121970A1
; GENERAL INFORMATION:
; APPLICANT: Watkins, Jeffrey D.
; APPLICANT: Hancock, James D.
; TITLE OF INVENTION: Butyrylcholinesterase Variant
; TITLE OF INVENTION: Polypeptides with Increased Catalytic Efficiency and Methods
; TITLE OF INVENTION: of Use
; FILE REFERENCE: P-IX 5555
; CURRENT APPLICATION NUMBER: US/10/324,466
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 10/032,233
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Butyrylcholinesterase variant
; NAME/KEY: CDS
; LOCATION: (214)...(1935)
US-10-324-466-39

Alignment Scores:
Pred. No.: 4-62e-44 Length: 2416
Score: 456.00 Matches: 160
Percent Similarity: 45.1% Conservatives: 83
Best Local Similarity: 29.74% Mismatches: 185
Query Match: 15.92% Indels: 110
DB: 17 Gaps: 26

US-09-943-857-4 (1-547) x US-10-324-466-39 (1-2416)
QY 22 GlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGluAlaPheLeuGlyIleProPhe 41
Db 277 GGTGGCAGCGTAACA-----GCCTTTCTTGAATTCCTAT 312
QY 42 AlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySerLeu 61
Db 313 GCACGCCACCTCTGGTAGACTTCGATTCAAAAGCCACAGTCTCTGACCAAGTGGTCT 372
QY 62 AsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGlyThr 81
Db 373 GATATTGGAAATGCCCAAAATATGCAAAATTTCTTGCTGT---CAGAACATAGATCAAGT 429
QY 82 PheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101
Db 430 TTTCCAGGCTTCCATGGA-----TCAGAGATGTGGAC 462
QY 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro---ProGly 120
Db 463 CCAACACACTGACCTCAGTGAAGACTGTTTATATCTAAATGTATGATTCAGACACCTAAA 522
QY 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIle 140
Db 523 CCAAAA-----AATGCCACTGATGATGATGATGATGATGATGATGATGATGATGAT 573
QY 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160

```

```

Db 574 GGAACATCACTTTA-----CAATGTTTATGATGGCAAG 606
Qy 161 -----HisIleHisValAlaValAsnTyrArgValAlaSerTrp 174
Db 607 TTTCTGCTGGGTTGAAGAGTTATTGTAGTGTCAATGACTATAGGGTGGGCTTA 666
Qy 175 GlyPheLeuAla-----GlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeu 192
Db 667 GGATTTCTTAGCTTTGCCAGGAAT-----CCTGAGGCTCCAGGGAACATGGGTTTA 717
Qy 193 LysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyAsp 212
Db 718 TTTGATCAACAGTTGGCTCTTCAGTGGTTCAAAAAATATAGCAGCTTTGGTGGAAAT 777
Qy 213 ProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeu 232
Db 778 CCTAAAAGTGTACTCTCTTTGGAGAAAGTCAGAGCAGCTTCAGTTAGCTGCATTGG 837
Qy 233 IleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMet 252
Db 838 CTT---TCTCTGAGACCACTTA-----TTGTTCCACGAGCCATTCTG 879
Qy 253 GlnSerGlyAlaMetValProSerAspProValAspGlyThrTyr-----GlyAsnGlu 270
Db 880 CAAAGTGGATCCATGAATGCTCCTTGGGCGGTAACTCTCTTTATGAAGCTAGGAACAGA 939
Qy 271 IleTyrAspLeuPheValSerSerAlaGlyCysGly-----SerAlaSerAspLysLeu 288
Db 940 ACCTTGAACCTTA---GCTAAATTCAGCTGGTGTCTAGAGAGATGAGACTGAATAATC 996
Qy 289 AlaCysLeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGly 308
Db 997 AAGTGTCTTAGAAATAAGATCCCAAGAAATCTT-----CTGAATGAAGCATTT 1047
Qy 309 PheLeuAlaTyrSerSer---LeuArgLeuSerTyrLeuProArgProAspGlyLysAsn 327
Db 1048 GTTGTCCTCTATGGGACTCTTTGTTCAGTAAACTTTTGGTCCGCGGTGATGGTATTT 1107
Qy 328 IleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIle 347
Db 1108 CTCACCTGACATGCCAGACATATTACTTGAACCTTGGCAATTTAAAAACCCAGATTG 1167
Qy 348 IleGlyAspGlnAsnAspGluGlyThrIlePhe-----GlyLeuSer 361
Db 1168 GTGGGTGTTAATAAGATGAAGGAGCAGCTTTTATTAGTCTATGCTCTCTGCTTCAGC 1227
Qy 362 SerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPhe 379
Db 1228 AAAGATACAATAGTATCATCTAGAAAGAAATTTTCAGGAAGGTTTAAAAATATTTT 1287
Qy 380 IleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr-----394
Db 1288 CCAGGAGTGAGTGAGTTGGAAAGGAATCCATCTTTTCATTACACAGACTGGGTAGAT 1347
Qy 395 -----ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAla 411
Db 1348 GATCAGACACCTTGAAACTACCGTCAGGCTTGGGTGATGTTGTTGGGATTATAATTC 1407
Qy 412 LeuThrPro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIle 429
Db 1408 ATATGCCCTGCTTGGAGTTTCCACCAAGAGTTCCTCAGAAATGGGGAATAATGCTTTTTC 1467
Qy 430 HisAlaArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSer 449
Db 1468 -----TACTATTTCAGAC-----CGATCC 1488
Qy 450 LysGlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGln 469
Db 1489 TCCAAACTTCCGTGGCCGAATGGATGGAGTGATGCTGATGCTATCAAT-----1539
Qy 470 AspTyrLeuLeuGly-----474

```

```

Db 1540 GAATTTCTGTTGTTTACCTCTGAAAGAAGAGATAAATTACACAAAAGCCGAGAAATT 1599
Qy 475 ---SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspPro 493
Db 1600 TTGAGTAGATCCATAGTG-----AAACGGTGGGCAAAATTTTGCAAAATATGGGAATCCA 1653
Qy 494 Asn---ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db 1654 AATGAGACTCAGAACATAGCACAAGCTGGCTCTCTTCAAAAGCACTGAACAA 1707

RESULT 9
US-10-413-432-39
; Sequence 39, Application US/10413432
; Publication No. US20040120939A1
; GENERAL INFORMATION:
; APPLICANT: Watkins, Jeffrey D.
; APPLICANT: Hancock, James D.
; TITLE OF INVENTION: Butyrylcholinesterase Variant
; TITLE OF INVENTION: Polypeptides with Increased Catalytic Efficiency and Methods
; FILE REFERENCE: P-IX 5510
; CURRENT APPLICATION NUMBER: US/10/413,432
; CURRENT FILING DATE: 2003-04-11
; PRIOR APPLICATION NUMBER: US 10/324,466
; PRIOR FILING DATE: 2002-12-20
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Butyrylcholinesterase variant
; NAME/KEY: CDS
; LOCATION: (214)...(1935)
US-10-413-432-39

Alignment Scores:
Pred. No.: 4,62e-44 Length: 2416
Score: 456.00 Matches: 160
Percent Similarity: 45.17% Conservative: 83
Best Local Similarity: 29.74% Mismatches: 185
Query Match: 15,92% Indels: 110
DB: 17 Gaps: 26

US-09-943-857-4 (1-547) x US-10-413-432-39 (1-2416)
Qy 22 GlyAspThrIleThrGlyLeuAsnAlaIleIleAsnGluAlaPheLeuGlyIleProPhe 41
Db 277 GGTGGCAGCGTAACA-----GCCTTTCTTGGAAATTCCTAT 312
Qy 42 AlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySerLeu 61
Db 313 GCACAGCCACCTCTTGTAGACTTCGATTCAAAAAGCCACAGCTCTCTGACCAAGTGTCT 372
Qy 62 AsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGlyThr 81
Db 373 GATATTTGGAAATGCCACAAATATATCAAAATTCCTCTGT---CAGAACATATGCAAACT 429
Qy 82 PheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101
Db 430 TTTCCAGGCTTCCATGGA-----TCAGATGATGGAAAC 462
Qy 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro---ProGly 120
Db 463 CCAAACTGACCTCAGTCAAGACTGTTTATATCTAAATGATGATTCACACACCTAAA 522
Qy 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIle 140
Db 523 CCAAAA-----AATGCCACTGTATTGATGATTTATGTTGGTGGTGGTTTCAAACT 573
Qy 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160

```

1540 GAAATTTGCTTTTGGTTTACCTCTCGGAAAGAGAGATAATTACAAAGCCGAGGAAT 1591  
 475 ---SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspPro 493  
 1600 TTGAGTAGATCATAGTG-----AAACGGTGGCBAATTTTGCAAAATATGGGAATCCA 1653  
 494 Asn---ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerGln 510  
 1654 AATGAGACTCAGAACATAGCAAGAGCTGGCTCTCTTCAAAAGCACTGAACAA 1707  
 RESULT 10  
 US-09-748-739A-7  
 ; Sequence 7, Application US/09748739A  
 ; Patent No. US20020119489A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lockridge, Oksana  
 ; APPLICANT: Watkins, Jeffrey D.  
 ; TITLE OF INVENTION: Butyrylcholinesterase Variants and  
 ; TITLE OF INVENTION: Methods of Use  
 ; FILE REFERENCE: P-IX 4143  
 ; CURRENT APPLICATION NUMBER: US/09/748,739A  
 ; CURRENT FILING DATE: 2000-12-06  
 ; NUMBER OF SEQ ID NOS: 31  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 7  
 ; LENGTH: 2416  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Human Butyrylcholinesterase variant  
 ; NAME/KEY: CDS  
 ; LOCATION: (214)...(1935)  
 US-09-748-739A-7  
 Alignment Scores:  
 Pred. No.: 6,1e-44 Length: 2416  
 Score: 455.00 Matches: 160  
 Percent Similarity: 45.17% Conservative: 83  
 Best Local Similarity: 29.74% Mismatches: 195  
 Query Match: 15.89% Indels: 110  
 DB: 9 Gaps: 26  
 US-09-943-857-4 (1-547) x US-09-748-739A-7 (1-2416)  
 22 GlyAspThrIleThrGlyLeuAsnAlaIleAsnGluAlaPheLeuGlyIleProPhe 41  
 277 GTGGCAAGGTAAACA-----GCCTTTCTTGGAAATCCCTAT 312  
 42 AlaGluProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySerLeu 61  
 313 GCACAGCACCTCTGTGTAGACTTCGATTCAAAAGCCACAGTCTCTGACCAAGTGGTCT 372  
 62 AsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGlyThr 81  
 373 GATATTTGGAAATGCCACAAATATGCAATTTCTTGTGTG---CAGAACATAGTCAAAGT 429  
 82 PheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101  
 430 TTTCCAGGCTTCCATGGA-----TCAGAGATGTGGAAAC 462  
 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro---ProGly 120  
 463 CCAACACACTGACCTCAGTGAAGACTGTATATATCTAAATGTATGGATTCACAGACCTTAA 522  
 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIle 140  
 523 CCAAAA-----ATGCCACTGATTGATGATGATTTATGTTGGTGGTGGTTCCTCAAACT 573  
 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160  
 574 GGAACATCATCTTTA-----CATGTTATGATGTCGAAG 606

QY 161 -----HisIleHisValalaValaAsnTyrArgValalaSerTrp 174  
 Db 607 TTTCGTGCTGGGTGAAGAGTTATTGTAGTCAATGAATATAGGTCGGCTA 666  
 QY 175 GlyPheLeuAla-----GlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeu 192  
 Db 667 GGATTCCTTAGCTTGCAGGAAAT-----CCTGAGGCTCCAGGGAAACATGGGTTA 717  
 QY 193 LysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAsp 212  
 Db 718 TTTCATCAACAGTTGGCTCTTCAGTGGTTCAAAAAATATAGACGCTTTGGTGGAAAT 777  
 QY 213 ProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuGlyHisLeu 232  
 Db 778 CCTAAAGTGTAACTCTCTTTGGAGAAAGTCAGAGCAGCTTCAGTTAGCTGCATTG 837  
 QY 233 IleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMet 252  
 Db 838 CTT---TCTCTGGAACCAITCA-----TTGTTCCACAGAGCCATTCTG 879  
 QY 253 GlnSerGlyAlaMetValProSerAspProValAspGlyThrTyr-----GlyAsnGlu 270  
 Db 880 CAAAGTGCATCTTAATGCTCTTGGCGGTAAACATCTCTTATGAGCTAGAGACAGA 939  
 QY 271 IleTyrAspLeuPheValSerAlaGlyCysGly-----SerAlaSerAspLysLeu 288  
 Db 940 ACCTGTGAACCTA---GCTAAATTGACTGTGTCTAGAGAGATGAGACTGAATAATC 996  
 QY 289 AlaCysLeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGly 308  
 Db 997 AAGTGTCTTGAATAAAGATCCCAAGAAATCTT-----CTGAATGAAGCAATTT 1047  
 QY 309 PheLeuAlaTyr---SerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsn 327  
 Db 1048 GTGTGCTCCCTATGGGACTGTGTGTCAGTAACTTTGGTCGACCGGTGGATGATTT 1107  
 QY 328 IleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIle 347  
 Db 1108 CTCACCTGACATGCCAGACATATTACTTGAACCTTGGACAAATTAATAAAACCCAGATTTG 1167  
 QY 348 IleGlyAspGlnAsnAspGluGlyThrIlePhe-----GlyLeuSer 361  
 Db 1168 GTGGGTGTTAAAGATGAGGAGCAGCTTTTATGCTATGCTGCTCTGCTTCAGC 1227  
 QY 362 SerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPhe 379  
 Db 1228 AAAGATAACAATAGTATCATCACTAGAAAGAAATTTTCAGGAAGTTTAAATAATTTTT 1287  
 QY 380 IleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr----- 394  
 Db 1288 CCAGGAGTGAAGTGGTTGGAAAGGAATCCATCTTTTTCATTACACAGACTGGTAGAT 1347  
 QY 395 -----ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeuAsnAla 411  
 Db 1348 GATCAGACCTGAAACCTACCGTAGGCTTGGGTGATGTTGTTGGGATTAATAATTTTC 1407  
 QY 412 LeuThrPro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIle 429  
 Db 1408 ATATGCCCTGCTTGGAGTTCACCAAGAGTCTCAGAAATGGGAATAATGCTTTTTC 1467  
 QY 430 HisAlaArgArgTyrPheLeuAsnHisPheGlnGlyGlyThrLysTyrSerPheLeuSer 449  
 Db 1468 -----TACTATTTTGAACAC-----CGATCC 1488  
 QY 450 LysGlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGln 469  
 Db 1489 TCCAAACTTCCGTGGCCAGAAATGGATGGGATGATGATGGCTATGAAAT----- 1539  
 QY 470 AspTyrLeuLeuGly----- 474  
 Db 1540 GAATTGTCTTTGGTTTACCTCTGGAAGAGAGATAATTACAAAAACCCGAGGAATTT 1599  
 QY 475 ---SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspPro 493

Db 1600 TTGAGTAGATCCATAGTG-----AAACGGTGGCAAAATTTTGCATAATATGGGAATCCA 1653  
 QY 494 Asn---ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510  
 Db 1654 AATGAGACTCAGAAACAATAGCAAAAGCTGGCTGTCTTCAAAGACACTGAACAA 1707  
 RESULT 11  
 US-09-810-861B-5  
 ; Sequence 5, Application US/09810861B  
 ; Patent No. US20020162140A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mor, Tsafir S.  
 ; APPLICANT: Soreq, Hermona  
 ; APPLICANT: Arntzen, Charles J.  
 ; APPLICANT: Maenon, Hugh S.  
 ; TITLE OF INVENTION: EXPRESSION OF RECOMBINANT HUMAN ACETYLCHOLINESTERASE IN  
 ; FILE REFERENCE: BTI-45  
 ; CURRENT APPLICATION NUMBER: US/09/810,861B  
 ; PRIOR FILING DATE: 2001-03-16  
 ; PRIOR APPLICATION NUMBER: 60/190,440  
 ; PRIOR FILING DATE: 2000-03-17  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: PatentIn Ver. 3.1  
 ; SEQ ID NO 5  
 ; LENGTH: 1725  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
 ; OTHER INFORMATION: human acetylcholinesterase gene optimized for  
 ; OTHER INFORMATION: expression in plants  
 US-09-810-861B-5  
 Alignment Scores:  
 Pred. No.: 8,21e-44 Length: 1725  
 Score: 452.00 Matches: 156  
 Percent Similarity: 44.59% Conservative: 83  
 Best Local Similarity: 29.10% Mismatches: 179  
 Query Match: 15.78% Indels: 118  
 DB: Gaps: 26  
 US-09-943-857-4 (1-547) x US-09-810-861B-5 (1-1725)  
 QY 35 AlaPheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspPro 54  
 Db 184 GCTTCTCTGGCATCCCTTTGCGAGACCAACCCATGGGACCCCGTCGCTTCTGCCACCG 243  
 QY 55 -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71  
 Db 244 GAGCCCAACGACGCTTGGTCAGGGGTGTAGACGCTTACACCTTCCAGAGTGTCTC 297  
 QY 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85  
 Db 298 ---TGCTACCATATGTGGACACCCCTATACCCAGGTTTGGAGGACACCGAGATGTGAAC 354  
 QY 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105  
 Db 355 CCCAACCCGTGAGCTG----- 369  
 QY 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125  
 Db 370 ---AGCAGGAGCTGCTGTACCTCAAC---GTGTGGACACACCATACCCCGGCTATATCC 423  
 QY 126 AsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIleGlySerProThrIle 145  
 Db 424 CCCACCCCTGCTCTGCTGTGATCTATGGGGTGGCTTCTACAGTGGGGCTCTCTCTTG 483  
 QY 146 PheProProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleHis--- 164  
 Db 484 -----GACGTGTACGATGGCCGCTTCTTGTGTACAGGCC 516

```

165 QY -----ValAlaValAsnTyrArgValAlaSerTyrGlyPheLeuAlaGly 179
166 QY |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 Db GAGAGGACTGTGGTGTGCATCACTACCGGTGGGACCTTGGCTTCTTGGCCCTG 576
168 QY 180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199
169 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
170 QY 577 CCGGGGAGCGA---GAGGCGCGGCAATGTGGTCTCTGGATCAGAGGCTGGCCCTG 633
171 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 QY 200 GlnTyrValAlaAspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePhe 219
173 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 QY 634 CAGTGGGTGAGGAGAGAGCTGGGAGCTTCCGAGGCTTCCGAGCATCAGTGCAGCTGTTT 693
175 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
176 QY 220 GlyLysSerAlaGlySerMetSerValLeuCysHisLeuIleTyrAsnAspGlyAspAsn 239
177 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
178 QY 694 GGGGAGAGCGGGAGCGGCTCGTGGGCGATGCACCTGCTG-----735
179 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
180 QY 240 ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
181 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
182 QY 736 TCCCGCGGAGCGGGGCTGTCCACAGGCGCGTGTGCAGAGCGGTGCC-----CCC 789
183 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 QY 260 SerAspProValAspGlyThrTyrGly-----AsnGluIle 271
185 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 QY 790 AATGACCCCTGG---GCCACGCTGGGCAATGGGAGAGCGCCGTCGAGGCGCCACGAGCTG 846
187 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 QY 272 TyrAspLeuPheValSerAlaGlyCys-----GlySerAlaSer 285
189 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
190 QY 847 GCCCACCCTT-----GTGGCTGTCTCTCCAGCGGCGACCTGGTGGGAATCAGACA 894
191 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192 QY 286 AspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn 305
193 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 QY 895 GAGCTGTAGCTGCTCGGACAGCAGCAGCGAGGCTCTGGT-----AACAC 945
195 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 QY 306 ThrProGlyPheLeuAlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgPro 324
197 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 QY 946 GAATGGCAGCTGCTCCCTCAGAAAGCGTCTTCCGGTCTCTCTCGTGGTGTAGAT 1005
199 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 QY 325 GlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerVal 344
201 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 QY 1006 GGAGACTTCTCAGTGAACACCCAGAGGCGCTTCATCAGCGGAGAGCTTCCACGCGCTG 1065
203 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 QY 345 ProValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe-----358
205 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206 QY 1066 CAGGTGCTGTGGTGTGTGAAGGATGAGGCTCTGATTTTCTGTTTACGGGCGCCCA 1125
207 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
208 QY 359 GlyLeuSerSerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLys 376
209 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 QY 1126 GGCCTTCAGCAAGACAAACAGTCTCTCATCAGCGGCGCGAGTTCTCGCGGGGTGGG 1185
211 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
212 QY 377 GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr-----394
213 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 QY 1186 GTCGGGGTTCGCCAGTAAGTACCTGGCGGCGGAGGCTGTGGTCTGCATTACACAGAC 1245
215 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
216 QY 395 -----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal 408
217 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218 QY 1246 TGGCTGCATCCCGAGGACCGGCGACGCTGTGAGGAGCGGCTGAGCGATGTGTGGCGAG 1305
219 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 QY 409 LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe 428
221 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 QY 1306 CACAATGTGTGTGCGCC-----GTGGCCAGTGTGCTGGGCGAGTGGTGGCC 1353
223 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
224 QY 429 IleHisAlaArg-----ArgTyrPheLeuAsnHisPheGlnGly-----441
225 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
226 QY 1354 CAGGTGCGCGGTCTACGCTTGTGAACACCGTCTTCCAGGCTCTCTCGGCC 1413
227 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
228 QY 442 -----GlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIle---457
229 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
230 QY 1414 CTGTGATGGGGTGGCCCGGCTACGAGATCAGTTCATCTTTGGGATCCCGCTGGAC 1473
231 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
232 QY 458 ---MetGlyThrPheHisAlaAsnAspIleValTyrGlnAspTyrLeuLeuGlySerGly 476
233 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 QY 1474 CCTCTCGAAACTACGCGGAGGAGAAATCTTCGCCAGGCGACTGATG-----1524
235 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 QY 477 SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496
237 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1525 Db -----CGAFACTGGGCCAACTTTGCCGCGACACGGGATCCCAATGAGCCC 1569
1526 QY 497 -----GlyLeuLeuValAsnTyrProLysTyrThrSerSerSerGln 510
1527 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1528 QY 1570 CGAGACCCCAAGCCCCCAATGCCCCCGTACACGCGGGGGTCTCAG 1617
1529 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1530 RESULT 12
1531 US-10-032-233-33
1532 ; Sequence 33, Application US/10032233
1533 ; Publication No. US20030153062A1
1534 ; GENERAL INFORMATION:
1535 ; APPLICANT: Watkins, Jeffrey D.
1536 ; APPLICANT: Hancock, James D.
1537 ; TITLE OF INVENTION: Butyrylcholinesterase Variants with
1538 ; TITLE OF INVENTION: Increased Catalytic Efficiency and Methods of Use
1539 ; FILE REFERENCE: P-IX 4642
1540 ; CURRENT APPLICATION NUMBER: US/10/032,233
1541 ; CURRENT FILING DATE: 2001-12-20
1542 ; NUMBER OF SEQ ID NOS: 50
1543 ; SOFTWARE: FastSeq for Windows Version 4.0
1544 ; SEQ ID NO 33
1545 ; LENGTH: 2416
1546 ; TYPE: DNA
1547 ; ORGANISM: Artificial Sequence
1548 ; FEATURE:
1549 ; OTHER INFORMATION: Butyrylcholinesterase variant
1550 ; NAME/KEY: CDS
1551 ; LOCATION: (214)...(1935)
1552 ; US-10-032-233-33
1553 Alignment Scores:
1554 Pred. No.: 1,41e-43 Length: 2416
1555 Score: 452.00 Matches: 161
1556 Percent Similarity: 45.37% Conservative: 84
1557 Best Local Similarity: 29.81% Mismatches: 181
1558 Query Match: 15.78% Indels: 114
1559 DB: 15 Gaps: 28
1560
1561 US-09-943-857-4 (1-547) x US-10-032-233-33 (1-2416)
1562 QY 22 GlyAspThrIleThrGlyLeuAsnAlaIlelleAsnGluAlaPheLeuGlyIleProPhe 41
1563 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1564 QY 277 GGTGCGACGGTAACA-----GCCTTCTTGGATTCCTAT 312
1565 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1566 QY 42 AlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySerLeu 51
1567 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1568 QY 313 GCACAGCCACCTCTTGGTAGACTTCGATTCAAAAAGCCACAGTCTCTGACCAAGTGGTCT 372
1569 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1570 QY 62 AsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGlyThr 81
1571 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1572 QY 373 GATATTGGAATGCCACAAAATATGCAATTCCTGCTGT---CAGAACATAGATCAAACT 429
1573 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1574 QY 82 PheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101
1575 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1576 QY 430 TTTCAGGCTTCCATGGA-----TCAGATGTGGAAC 462
1577 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1578 QY 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro---ProGly 120
1579 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1580 QY 463 CCAACACATGACTCAGTCAAGACTGTTTATATCATTAATGTATGATTCAGCACCCTAA 522
1581 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1582 QY 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuThrIlePheGlyGlyGlyIle 140
1583 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1584 QY 523 CCAAAA-----AATGCCACTGTATTGATGATGATTTATGTTGGTGGTTCAACT 573
1585 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1586 QY 141 GlySerProThrIlePheProProAlaGlnMetValThrLysSerValLeuMetGlyLys 160
1587 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1588 QY 574 GGAACATCATCTTA-----CATGTTTATGATGGAAC 606
1589 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1590 QY 161 -----HisIleIleHisValAlaValAsnTyrArgValAlaSerTyr 174
1591 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1592 QY 607 TTCTCGCTCGGTGTTGAAGAGTATTATTAGTGTCAATGAACATATAGGGTGGGTGCCCTA 665
1593 Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```

QY 175 GlyPheLeuAla-----GlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeu 192
Db 567 GGATCTTAGCTTTGCCAGAAAT-----CCTGAGGCTCCAGGGAACATGGTTTA 717

QY 193 LysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAsp 212
Db 718 TTTCATCAACAGTTGGCTCTTCAGTGGTTCAAAAAATATAGCAGCCTTTGGTGGAAAT 777

QY 213 ProSerLysValThrIlePheGlyGlySerAlaGlySerMetSerValLeuLysCysHisLeu 232
Db 778 CCTAAAAGTGTAACTCTCTTTGGAGAAAGTGCAGGAGCAGCTTCAGTTAGCTGCATTG 837

QY 233 IleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyLeuMet 252
Db 838 CTT---TCTCTGGAACCCATCA-----TTGTTCCAGCAGCCATCTGT 879

QY 253 GlnSerGlyAlaMetValProSerAspPro-----ValAspGlyThrTyr-----Gly 268
Db 880 CAAAGTGGATCC-----CCGAATGCTCTTGGCGGTAAACATCTTTATGAAAGCTAGG 933

QY 269 AsnGluIleTyrAspLeuPheValSerSerAlaGlyCysGly-----SerAlaSerAsp 286
Db 934 AACAGAACGTTGAACCTTA---GCTAAATTCAGCTGTTGCTCTAGAGAGATGACACTGAA 990

QY 287 LysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThr 306
Db 991 ATAAATCAAGTGTCTTGAATAAAGATCCCCAAGAAATCTT-----CTGAATGAA 1041

QY 307 ProGlyPheLeuAlaTyrSerSer---LeuArgLeuSerTyrLeuProArgProAspGly 325
Db 1042 GCATTGTGTCCTCCATGGAGCTCTCTTGTGATAAATCTTGTGTCGACCGGTGGATGGT 1101

QY 326 LysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValPro 345
Db 1102 GATTTCTCACTGACATGCCAGACATATTACTTGAACCTTGACACATTTAAAAAACCCAG 1161

QY 346 ValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe-----Gly 359
Db 1162 ATTTTGGTGGTGTAAATAAAGATGAAGGACGCTTTTTTACTGTATGGTGTCTCTGGC 1221

QY 360 LeuSerSerLeuAsn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGln 377
Db 1222 TTCAGCAAGATCAACATAGTATCATCACTAGAAAAGAAATTCAGGAAGGTTTAAAAATA 1281

QY 378 SerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaIleTyr-----394
Db 1282 TTTTTCAGGAGTGTAGTGAGTTTGGAAAGGAATCCATCTTTTTCATTACACAGACTGG 1341

QY 395 -----ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeu 409
Db 1342 GTAGATGATCAGAGACCTGAAACCTACCGTGAGCCCTTGGGTGATGTTGTTGGGATTAT 1401

QY 410 AsnAlaLeuThrPro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAla 427
Db 1402 AATTTTCATATGCTCTGCTTGGAGTTCCACAGAGAGTTCTCAGAAATGGGAAATAATGCC 1461

QY 428 PheIleHisAlaArgArgTyrPheLeuAsnHisPheGlnGlyThrIleTyrSerPhe 447
Db 1462 TTTTTC-----TACTATTTTGAACAC-----1482

QY 448 LeuSerLysGlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleVal 467
Db 1483 CGATCTCCCAACTCTCCGTCGCGCAGAAATGGATGGAGTGATGATGCTATGAAATTT---1539

QY 468 TrpGlnAspTyrLeuLeuGly-----1540
Db 1540 -----GAATTTGCTTTGGTTTACCTCTGGAAGAGAGATAATTACAAAAACCCGAG 1593

QY 475 -----SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeu 491
Db 1594 GAAATTTTGTAGTAGTATCCATAGT-----AAACGCTGGGCAAAATTTTGCAAAATATGG 1647

QY 492 AspProAsn---ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerGln 510

```

```

Db 1648 AATCCAAATGAGACTCAGAACATAGCAAGCTGGCTGTCTTCAAAGGCACTGACAA 1707

RESULT 13
US-10-324-466-33
; Sequence 33, Application US/10324466
; Publication No. US20040121970A1
; GENERAL INFORMATION:
; APPLICANT: Watkins, Jeffrey D.
; APPLICANT: Hancock, James D.
; TITLE OF INVENTION: Butyrylcholinesterase Variant
; TITLE OF INVENTION: Polypeptides with Increased Catalytic Efficiency and Methods
; FILE REFERENCE: P-IX 5555
; CURRENT APPLICATION NUMBER: US/10/324,466
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 10/032,233
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Butyrylcholinesterase variant
; NAME/KEY: CDS
; LOCATION: (214) ... (1935)
US-10-324-466-33

```

## Alignment Scores:

```

Pred. No.: 1,41e-43 Length: 2416
Score: 452.00 Matches: 161
Percent Similarity: 45.37% Conservative: 84
Best Local Similarity: 29.81% Mismatches: 181
Query Match: 15.78% Indels: 114
DB: 17 Gaps: 28

```

US-09-943-857-4 (1-547) x US-10-324-466-33 (1-2416)

```

QY 22 GlyAspThrIleThrGlyLeuAsnAlaIleAsnGluAlaPheLeuGlyIleProPhe 41
Db 277 GGTGGCAGGTAAACA-----GCCTTCTTGAATTCCTAT 312

QY 42 AlaGluProProValGlyAsnLeuArgPheLysAspProValProTyrSerGlySerLeu 61
Db 313 GCACAGCCACCTCTTGTGTAGACTTCGATTCAAAAAGCCAGCTCTGACCAAGTGTCT 372

QY 62 AsnGlyGlnLysPheThrSerTyrGlyProSerCysMetGlnGlnAsnProGluGlyThr 81
Db 373 GATATTGGAAATGCCCAAAATATGCAAAATCTTGTCTGT---CAGAACATAGATCAAAAGT 429

QY 82 PheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101
Db 430 TTTCCAGGCTTCCATGGA-----TCAGAGATGTGGAAC 462

QY 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro---ProGly 120
Db 463 CCAAAACACTGACCTCAGTGAAGACTTTATATCTAAATGTATGGATTCACGACCTAAA 522

QY 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIle 140
Db 523 CCAAAA-----AATGCCACTGTATGTATGATGATTTATGGTGGTGGTTTCAAACT 573

QY 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160
Db 574 GGAACATCATCTTAA-----CATGTTTATGATGCGCAAG 606

QY 161 -----HisIleIleHisValAlaValAsnTyrArgValAlaSerTrp 174
Db 607 TTTCTGGCTCGGTTGAAAGAGATTATTAGTGTCAATGAACATATAGGTGGTGGCTCA 666

```

```

175 GlyPheLeuLa-----GlyAspAspIleIleGluGlySerGlyAsnAlaGlyLeu 192
176 |||||
177 |||||
178 |||||
179 |||||
180 |||||
181 |||||
182 |||||
183 |||||
184 |||||
185 |||||
186 |||||
187 |||||
188 |||||
189 |||||
190 |||||
191 |||||
192 |||||
193 LysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAsp 212
194 |||||
195 |||||
196 |||||
197 |||||
198 |||||
199 |||||
200 |||||
201 |||||
202 |||||
203 |||||
204 |||||
205 |||||
206 |||||
207 |||||
208 |||||
209 |||||
210 |||||
211 |||||
212 |||||
213 ProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeu 232
214 |||||
215 |||||
216 |||||
217 |||||
218 |||||
219 |||||
220 |||||
221 |||||
222 |||||
223 |||||
224 |||||
225 |||||
226 |||||
227 |||||
228 |||||
229 |||||
230 |||||
231 |||||
232 |||||
233 IleTrpAsnAspGlyAspAsnThrTyrlsGlyLysProLeuPheArgAlaGlyIleMet 252
234 |||||
235 |||||
236 |||||
237 |||||
238 |||||
239 |||||
240 |||||
241 |||||
242 |||||
243 |||||
244 |||||
245 |||||
246 |||||
247 |||||
248 |||||
249 |||||
250 |||||
251 |||||
252 |||||
253 GlnSerGlyAlaMetValProSerAspPro-----ValAspGlyThrTyrls-----Gly 268
254 |||||
255 |||||
256 |||||
257 |||||
258 |||||
259 |||||
260 |||||
261 |||||
262 |||||
263 |||||
264 |||||
265 |||||
266 |||||
267 |||||
268 |||||
269 AsnGluIleTyAspLeuPheValSerSerAlaGlyCysGly-----SerAlaSerAsp 286
270 |||||
271 |||||
272 |||||
273 |||||
274 |||||
275 |||||
276 |||||
277 |||||
278 |||||
279 |||||
280 |||||
281 |||||
282 |||||
283 |||||
284 |||||
285 |||||
286 |||||
287 LysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThr 306
288 |||||
289 |||||
290 |||||
291 |||||
292 |||||
293 |||||
294 |||||
295 |||||
296 |||||
297 |||||
298 |||||
299 |||||
300 |||||
301 |||||
302 |||||
303 |||||
304 |||||
305 |||||
306 |||||
307 ProGlyPheLeuAlaTyrlsSerSer---LeuArgLeuSerTyrlsLeuProArgProAspGly 325
308 |||||
309 |||||
310 |||||
311 |||||
312 |||||
313 |||||
314 |||||
315 |||||
316 |||||
317 |||||
318 |||||
319 |||||
320 |||||
321 |||||
322 |||||
323 |||||
324 |||||
325 |||||
326 LysAsnIleThrAspAspMetTyrlsLeuValArgAspGlyTyrlsTyrlsAlaSerValPro 345
327 |||||
328 |||||
329 |||||
330 |||||
331 |||||
332 |||||
333 |||||
334 |||||
335 |||||
336 |||||
337 |||||
338 |||||
339 |||||
340 |||||
341 |||||
342 |||||
343 |||||
344 |||||
345 |||||
346 ValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe-----Gly 359
347 |||||
348 |||||
349 |||||
350 |||||
351 |||||
352 |||||
353 |||||
354 |||||
355 |||||
356 |||||
357 |||||
358 |||||
359 |||||
360 |||||
361 |||||
362 |||||
363 |||||
364 |||||
365 |||||
366 |||||
367 |||||
368 |||||
369 |||||
370 |||||
371 |||||
372 |||||
373 |||||
374 |||||
375 |||||
376 |||||
377 |||||
378 SerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrls----- 394
379 |||||
380 |||||
381 |||||
382 |||||
383 |||||
384 |||||
385 |||||
386 |||||
387 |||||
388 |||||
389 |||||
390 |||||
391 |||||
392 |||||
393 |||||
394 |||||
395 |||||
396 |||||
397 |||||
398 |||||
399 |||||
400 |||||
401 |||||
402 |||||
403 |||||
404 |||||
405 |||||
406 |||||
407 |||||
408 |||||
409 |||||
410 |||||
411 |||||
412 |||||
413 |||||
414 |||||
415 |||||
416 |||||
417 |||||
418 |||||
419 |||||
420 |||||
421 |||||
422 |||||
423 |||||
424 |||||
425 |||||
426 |||||
427 |||||
428 PheIleHisAlaArgArgTyrlsPheLeuAsnHisPheGlnGlyThrLysTyrlsSerPhe 447
429 |||||
430 |||||
431 |||||
432 |||||
433 |||||
434 |||||
435 |||||
436 |||||
437 |||||
438 |||||
439 |||||
440 |||||
441 |||||
442 |||||
443 |||||
444 |||||
445 |||||
446 |||||
447 |||||
448 LeuSerLysGlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleVal 467
449 |||||
450 |||||
451 |||||
452 |||||
453 |||||
454 |||||
455 |||||
456 |||||
457 |||||
458 |||||
459 |||||
460 |||||
461 |||||
462 |||||
463 |||||
464 |||||
465 |||||
466 |||||
467 |||||
468 TrpGlnAspTyrlsLeuLeuGly-----TACTATTGTGAACAC----- 474
469 |||||
470 |||||
471 |||||
472 |||||
473 |||||
474 |||||
475 |||||
476 |||||
477 |||||
478 |||||
479 |||||
480 |||||
481 |||||
482 |||||
483 |||||
484 |||||
485 |||||
486 |||||
487 |||||
488 |||||
489 |||||
490 |||||
491 |||||
492 |||||
493 |||||
494 |||||
495 |||||
496 |||||
497 |||||
498 |||||
499 |||||
500 |||||
501 |||||
502 |||||
503 |||||
504 |||||
505 |||||
506 |||||
507 |||||
508 |||||
509 |||||
510 |||||
511 |||||
512 |||||
513 |||||
514 |||||
515 |||||
516 |||||
517 |||||
518 |||||
519 |||||
520 |||||
521 |||||
522 |||||
523 |||||
524 |||||
525 |||||
526 |||||
527 |||||
528 |||||
529 |||||
530 |||||
531 |||||
532 |||||
533 |||||
534 |||||
535 |||||
536 |||||
537 |||||
538 |||||
539 |||||
540 |||||
541 |||||
542 |||||
543 |||||
544 |||||
545 |||||
546 |||||
547 |||||
548 |||||
549 |||||
550 |||||
551 |||||
552 |||||
553 |||||
554 |||||
555 |||||
556 |||||
557 |||||
558 |||||
559 |||||
560 |||||
561 |||||
562 |||||
563 |||||
564 |||||
565 |||||
566 |||||
567 |||||
568 |||||
569 |||||
570 |||||
571 |||||
572 |||||
573 |||||
574 |||||
575 |||||
576 |||||
577 |||||
578 |||||
579 |||||
580 |||||
581 |||||
582 |||||
583 |||||
584 |||||
585 |||||
586 |||||
587 |||||
588 |||||
589 |||||
590 |||||
591 |||||
592 |||||
593 |||||
594 |||||
595 |||||
596 |||||
597 |||||
598 |||||
599 |||||
600 |||||
601 |||||
602 |||||
603 |||||
604 |||||
605 |||||
606 |||||
607 |||||
608 |||||
609 |||||
610 |||||
611 |||||
612 |||||
613 |||||
614 |||||
615 |||||
616 |||||
617 |||||
618 |||||
619 |||||
620 |||||
621 |||||
622 |||||
623 |||||
624 |||||
625 |||||
626 |||||
627 |||||
628 |||||
629 |||||
630 |||||
631 |||||
632 |||||
633 |||||
634 |||||
635 |||||
636 |||||
637 |||||
638 |||||
639 |||||
640 |||||
641 |||||
642 |||||
643 |||||
644 |||||
645 |||||
646 |||||
647 |||||
648 |||||
649 |||||
650 |||||
651 |||||
652 |||||
653 |||||
654 |||||
655 |||||
656 |||||
657 |||||
658 |||||
659 |||||
660 |||||
661 |||||
662 |||||
663 |||||
664 |||||
665 |||||
666 |||||
667 |||||
668 |||||
669 |||||
670 |||||
671 |||||
672 |||||
673 |||||
674 |||||
675 |||||
676 |||||
677 |||||
678 |||||
679 |||||
680 |||||
681 |||||
682 |||||
683 |||||
684 |||||
685 |||||
686 |||||
687 |||||
688 |||||
689 |||||
690 |||||
691 |||||
692 |||||
693 |||||
694 |||||
695 |||||
696 |||||
697 |||||
698 |||||
699 |||||
700 |||||
701 |||||
702 |||||
703 |||||
704 |||||
705 |||||
706 |||||
707 |||||
708 |||||
709 |||||
710 |||||
711 |||||
712 |||||
713 |||||
714 |||||
715 |||||
716 |||||
717 |||||
718 |||||
719 |||||
720 |||||
721 |||||
722 |||||
723 |||||
724 |||||
725 |||||
726 |||||
727 |||||
728 |||||
729 |||||
730 |||||
731 |||||
732 |||||
733 |||||
734 |||||
735 |||||
736 |||||
737 |||||
738 |||||
739 |||||
740 |||||
741 |||||
742 |||||
743 |||||
744 |||||
745 |||||
746 |||||
747 |||||
748 |||||
749 |||||
750 |||||
751 |||||
752 |||||
753 |||||
754 |||||
755 |||||
756 |||||
757 |||||
758 |||||
759 |||||
760 |||||
761 |||||
762 |||||
763 |||||
764 |||||
765 |||||
766 |||||
767 |||||
768 |||||
769 |||||
770 |||||
771 |||||
772 |||||
773 |||||
774 |||||
775 |||||
776 |||||
777 |||||
778 |||||
779 |||||
780 |||||
781 |||||
782 |||||
783 |||||
784 |||||
785 |||||
786 |||||
787 |||||
788 |||||
789 |||||
790 |||||
791 |||||
792 |||||
793 |||||
794 |||||
795 |||||
796 |||||
797 |||||
798 |||||
799 |||||
800 |||||
801 |||||
802 |||||
803 |||||
804 |||||
805 |||||
806 |||||
807 |||||
808 |||||
809 |||||
810 |||||
811 |||||
812 |||||
813 |||||
814 |||||
815 |||||
816 |||||
817 |||||
818 |||||
819 |||||
820 |||||
821 |||||
822 |||||
823 |||||
824 |||||
825 |||||
826 |||||
827 |||||
828 |||||
829 |||||
830 |||||
831 |||||
832 |||||
833 |||||
834 |||||
835 |||||
836 |||||
837 |||||
838 |||||
839 |||||
840 |||||
841 |||||
842 |||||
843 |||||
844 |||||
845 |||||
846 |||||
847 |||||
848 |||||
849 |||||
850 |||||
851 |||||
852 |||||
853 |||||
854 |||||
855 |||||
856 |||||
857 |||||
858 |||||
859 |||||
860 |||||
861 |||||
862 |||||
863 |||||
864 |||||
865 |||||
866 |||||
867 |||||
868 |||||
869 |||||
870 |||||
871 |||||
872 |||||
873 |||||
874 |||||
875 |||||
876 |||||
877 |||||
878 |||||
879 |||||
880 |||||
881 |||||
882 |||||
883 |||||
884 |||||
885 |||||
886 |||||
887 |||||
888 |||||
889 |||||
890 |||||
891 |||||
892 |||||
893 |||||
894 |||||
895 |||||
896 |||||
897 |||||
898 |||||
899 |||||
900 |||||
901 |||||
902 |||||
903 |||||
904 |||||
905 |||||
906 |||||
907 |||||
908 |||||
909 |||||
910 |||||
911 |||||
912 |||||
913 |||||
914 |||||
915 |||||
916 |||||
917 |||||
918 |||||
919 |||||
920 |||||
921 |||||
922 |||||
923 |||||
924 |||||
925 |||||
926 |||||
927 |||||
928 |||||
929 |||||
930 |||||
931 |||||
932 |||||
933 |||||
934 |||||
935 |||||
936 |||||
937 |||||
938 |||||
939 |||||
940 |||||
941 |||||
942 |||||
943 |||||
944 |||||
945 |||||
946 |||||
947 |||||
948 |||||
949 |||||
950 |||||
951 |||||
952 |||||
953 |||||
954 |||||
955 |||||
956 |||||
957 |||||
958 |||||
959 |||||
960 |||||
961 |||||
962 |||||
963 |||||
964 |||||
965 |||||
966 |||||
967 |||||
968 |||||
969 |||||
970 |||||
971 |||||
972 |||||
973 |||||
974 |||||
975 |||||
976 |||||
977 |||||
978 |||||
979 |||||
980 |||||
981 |||||
982 |||||
983 |||||
984 |||||
985 |||||
986 |||||
987 |||||
988 |||||
989 |||||
990 |||||
991 |||||
992 |||||
993 |||||
994 |||||
995 |||||
996 |||||
997 |||||
998 |||||
999 |||||
1000 |||||

```

QY 175 GlyPheLeuAala-----GlyAspPheLeuAalaGluGlySerGlyAsnAlaGlyLeu 192  
 Db 667 GATTCCTTAGCTTCCAGGAAT-----CCTGAGGCTCCAGGAACATGGGTTTA 717

QY 193 LysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyCysAsp 212  
 Db 718 TTTGATCAACAGTTGGCTCTTCAGTGGGTTTCAAAAAATATAGCAGCCTTTGGTGAAT 777

QY 213 ProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeu 232  
 Db 778 CCTAAAGTGTAACCTCTTTGGAGAAAGTCAGGAGCAGCTTCAGTTAGCTGCATTTG 837

QY 233 IleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAlaGlyIleMet 252  
 Db 838 CTT---TCTCTGGAAGCCATTCA-----TTGTTTCCAGCAGGCAATTCG 879

QY 253 GlnSerGlyAlaMetValProSerAspPro-----ValAspGlyThrTyr-----Gly 268  
 Db 880 CAAGTGGATCC-----CCGAATGCTCTTGGGGGGTAAACATCTCTTATGAAGCTAGG 933

QY 269 AsnGluIleTyrAspLeuPheValSerSerAlaGlyCysGly-----SerAlaSerAsp 286  
 Db 934 AACAGAACGTTGAACCTTA---GCTAAATTGACTGTTGCTCTAGAGAGATGAGACTGAA 990

QY 287 LysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThr 306  
 Db 991 ATATCAAGTGTCTTGAATAAAGATCCCAAGAAATTTCTT-----CTGAATGAA 1041

QY 307 ProGlyPheLeuAlaTyrSerSer---LeuArgLeuSerTyrLeuProArgProAspGly 325  
 Db 1042 GCATTTGTTGCTCCCTATGGACTCTTTGTCAGTAAACTTTGTCGCGACCGTGATGGT 1101

QY 326 LysAsnIleThrAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValPro 345  
 Db 1102 GATTTTCTCACTGACATGCGACATATATTCTGAACTTGACATTTTAAAAACCCAG 1161

QY 346 ValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe-----Gly 359  
 Db 1162 ATTTTGGTGGTGTATAAAGATGAAGGACAGCTTTTATGTCATGCTGCTCCTGGC 1221

QY 360 LeuSerSerLeuAsn-----ValThrAsnAlaGlnAlaArgAlaTyrPheLysGln 377  
 Db 1222 TTCAGCAAGATACATAGTATCATCACTAGAAAGAAATTCAGGAAGTTTAAAAATA 1281

QY 378 SerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr-----394  
 Db 1282 TTTTTCAGGAGTGAAGTGGTTGGAAGGAATCCATCCTTTTTCATACAGACTGG 1341

QY 395 -----ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeu 409  
 Db 1342 GTAGATGATCAGAGACTGAAACTACCGTGAGGCTTGGGTGATGTTGTTGGGATTA 1401

QY 410 AsnAlaLeuThrPro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAla 427  
 Db 1402 AATTTTCATATGCTCCTGCTGGAGTTCCACCAAGATTCACAGAATGGGGAATAATGCC 1461

QY 428 PheIleHisAlaArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPhe 447  
 Db 1462 TTTTTC-----TACTATTGTAACAC-----1482

QY 448 LeuSerLysGlnLeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleVal 467  
 Db 1483 CGATCTCCAAACTCCGTCGGCAGAAATGGATGGAGTGATGATGCTATGAAATT---1539

QY 468 TrpGlnAspTyrLeuLeuGly-----474  
 Db 1540 -----GAATTTGCTTTGGTTTACCTCTGGAAGAAAGAGATAATTACAAAAAGCCGAG 1593

QY 475 -----SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeu 491  
 Db 1594 GAAATTTTTCAGTAGATCCATAGT-----AAACGCTGGGCAAAATTTGCCAAATATGGG 1647

QY 492 AspProAsn---ThrAlaGlyLeuLeuValAsnTrpProLysTyrThrSerSerGln 510

Db 1648 AATCAATGAGACTCAGAACATAGCACAGCTGCTCTTCAAAAGCACTGAACAA 1707

RESULT 15  
 US-09-810-861B-3  
 ; Sequence 3, Application US/09810861B  
 ; Patent No. US20020162140A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mor, Tsafir S.  
 ; APPLICANT: Soreq, Hermona  
 ; APPLICANT: Arntzen, Charles J.  
 ; APPLICANT: Mason, Hugh S.  
 ; TITLE OF INVENTION: EXPRESSION OF RECOMBINANT HUMAN ACETYLCHOLINESTERASE IN  
 ; FILE REFERENCE: B71-45  
 ; CURRENT APPLICATION NUMBER: US/09/810,861B  
 ; CURRENT FILING DATE: 2001-03-16  
 ; PRIOR APPLICATION NUMBER: 60/190,440  
 ; PRIOR FILING DATE: 2000-03-17  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: PatentIn Ver. 3.1  
 ; SEQ ID NO 3  
 ; LENGTH: 5767  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: plasmid vector  
 ; OTHER INFORMATION: pTM034.  
 US-09-810-861B-3

Alignment Scores:  
 Pred. No.: 5,68e-43 Length: 5767  
 Score: 452.00 Matches: 156  
 Percent Similarity: 44.59% Conservative: 83  
 Best Local Similarity: 29.10% Mismatches: 179  
 Query Match: 15.78% Indels: 118  
 DB: 9 Gaps: 26

US-09-943-857-4 (1-547) x US-09-810-861B-3 (1-5767)

QY 35 AlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro 54  
 Db 1021 GCTTCTCTGGCATCCCTTTGGGAGCCACCCATGGAGCCCGTCGTTTCTGCCACCG 1080

QY 55 -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71  
 Db 1081 GAGCCCAAGCAGCCTTGGTCAGGGGTGTAGACGCTACACCTTCCAGAGTGTCTC-----1134

QY 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85  
 Db 1135 ---TGCTACCATATGTGGACACCCCTATACCCAGGTTTGGAGGCGCACCGAGATGTGGAAC 1191

QY 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105  
 Db 1192 CCCAACCGTGAGCTG-----1206

QY 106 GlnSerGluAspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAla 125  
 Db 1207 ---AGCGAGAGCTGCTGTACCTCAAC---GTGTGGACACCATACATCCCGGCTACATCC 1260

QY 126 AsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIleGlySerProThrIle 145  
 Db 1261 CCCACCCCTGCTCTGTCTGGAATCTATGGGGTGGCTTCTACAGTGGGCGCTCTCTCTTG 1320

QY 146 PheProProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHis---164  
 Db 1321 -----GACGTGTACGATGGCGCTTCTGTGTACAGGCC 1353

QY 165 -----ValAlaValAsnTyrArgValAlaSerTyrGlyPheLeuAlaGly 179  
 Db 1354 GAGAGAGCTGTCTGTGTCCATGAACCTACCGGTGGGAGCGCTTTGGTTCCTGCGCCCTG 1413

QY 180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199

Fri Aug 6 10:53:12 2004

us-09-943-857-4.rnpb

Db 2407 CGAGACCCCAAGGCCCCCAATGGCCCCCGTACACGGCGGGGCTCAG 2454

Search completed: August 5, 2004, 22:32:09  
Job time : 550 secs

```

1414 CCAGGGAGCCGA---GAGGCCCCGGGCAATGTGGTCTCTGGATCAGAGGCTGGCCCTG 1470
QY
200 GlnTyrValAlaAspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePhe 219
Db
1471 CAGTGGGTGAGGAGACGTGGCAGCCTTGGGGGTGACCCGACATCAGTGACGCTGTTT 1530
QY
220 GlyCysSerAlaGlySerMetSerValLeuLeuCysHisIleLeuTrpAsnAspGlyAspAsn 239
Db
1531 GGGGAGAGCGCGGAGCGCCTCGTGGGCATGACCTGCTG-----AsnGluIle 1572
QY
240 ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
Db
1573 TCCCGCCCGCCAGCGGGGCTGTTTCACAGGGCCGTGCTGCAGAGGGGTGCC-----CCC 1626
QY
260 SerAspProValAspGlyThrTyrGly-----AsnGluIle 271
Db
1627 AATGACCCCTGG---GCCACGCTGGCATGGAGAGGCCCGTCGCGAGGCCACGCGAGCTG 1683
QY
272 TyrAspLeuPheValSerSerAlaGlyCys-----GlySerAlaSer 285
Db
1684 GCCCACCTT-----GTGGGCTGTCTCCAGCGCGGCACTGGTGGGAATGACACA 1731
QY
286 AspLysLeuAlaCysLeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsn 305
Db
1732 GAGCTGGTAGCTGCTTCGGACACGACGACGCGCAGGTCTCTGGT-----AACCAC 1782
QY
306 ThrProGlyPheLeuAlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAsp 324
Db
1783 GAATGGCAGCTGCTGCCTCAAGAAAGCTCTTCGGGTCTCTCTGCTGCTGTGGTAGAT 1842
QY
325 GlyLysAsnIleThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerVal 344
Db
1843 GGAGACTTCTCAGTGACACCCCGAGGCCCTCATCAGCGGGGAGACTTCCACGGCCTG 1902
QY
345 ProValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe----- 358
Db
1903 CAGGTGCTGTGGTGGTGTGAGGATGAGGCTCGATTCTTGTGTTACGGGGCCCCA 1962
QY
359 GlyLeuSerSerLeuAsn-----ValThrAsnAlaGlnAlaArgAlaTyrPheLys 376
Db
1963 GGCTTCAGCAAAAGACAAACGAGTCTCTCATCAGCGGGCGGAGTTCCTGGCCGGGTGGG 2022
QY
377 GlnSerPheIleHisAlaSerAspAlaGluIleAspThrLeuMetAlaTyr----- 394
Db
2023 GTCGGGGTTCCCGAGGTAGTACCTGGCAGCGGCTGAGGCTGTGCTGCTGCTGCTACAGAC 2082
QY
395 -----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal 408
Db
2083 TGGCTGCATCCCGAGGACCCCGGACCGCTGAGGAGGCCCTGAGCGATGTGGTGGCGAC 2142
QY
409 LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe 428
Db
2143 CACAATGTCTGTGCCCC-----GTGGCCGAGCTGGCTGGGACTGGCTGCC 2190
QY
429 IleHisAlaArg-----ArgTyrPheLeuAsnHisPheGlnGly----- 441
Db
2191 CAGGGTGCCCGGGTCTACGCTTGTGAAACACCGCTGCTTCCAGGCTCTCTGGCCCC 2250
QY
442 -----GlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIle--- 457
Db
2251 CTGTGATGGGGTGGCCCGGACCGCTACGAGATCGAGTTCATCTTTGGGATCCCGCTGGAC 2310
QY
458 ---MetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuGlySerGly 476
Db
2311 CCCTCTCGAAACTACACGGCAGAGGAGAAATCTTCGCCGAGGACTGTG----- 2361
QY
477 SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496
Db
2362 -----CGAATCTGGGCCAACTTTGCCCGCACAGGGGATCCCATGAGCCC 2406
QY
497 -----GlyLeuLeuValAsnTrpProLysTyrThrSerSerGln 510

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 5, 2004, 18:59:17 ; Search time 2706 Seconds  
(without alignments)  
6036.442 Million cell updates/sec

Title: US-09-943-857-4  
Perfect score: 2864  
Sequence: 1 SNNRGPAGRLGSLVPTAKLA.....DNFRTAGYDALMTNPSSFFV 547

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-O=/cgn2\_1/USPTO/spool/US09943857/runat\_29072004\_090900\_29746/app\_query.fasta\_1.711  
-DB=EST -QFMT=fastap -SURFIX=rest -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09943857 @CGN 1 1 2607 @runat\_29072004\_090900\_29746 -NCPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPO2=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hcc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	586	20.5	769	14	CF881106	CF881106 trico83xn
c 2	586	20.5	824	14	CB907899	CB907899 trico83xn
c 3	554.5	19.4	719	14	CF880958	CF880958 trico83xc
c 4	554.5	19.4	774	14	CB907713	CB907713 trico83xc
c 5	536	18.7	970	14	CF885800	CF885800 trico83xn
c 6	481	16.8	1723	29	AY407738	AY407738 Mus muscu
c 7	481	16.8	2173	11	AK046080	AK046080 Mus muscu
c 8	481	16.8	2181	11	AK043748	AK043748 Mus muscu
c 9	481	16.8	2184	11	AK042335	AK042335 Mus muscu
c 10	471	16.4	3206	11	AK050337	AK050337 Mus muscu
c 11	467	16.3	2026	11	AK052200	AK052200 Mus muscu
c 12	461.5	16.1	2177	11	AK036259	AK036259 Mus muscu
c 13	461	16.1	620	10	AW790714	AW790714 D00142-R
c 14	456	15.9	604	14	CD056060	CD056060 H011A1S
c 15	452	15.8	1723	29	AY407736	AY407736 Homo sapi
c 16	452	15.8	1723	29	AY407737	AY407737 Pan trogl
c 17	452	15.8	1745	11	BC036813	BC036813 Homo sapi
c 18	452	15.8	2187	11	BC026315	BC026315 Homo sapi
c 19	451	15.7	2523	14	CD014089	CD014089 90134961
c 20	445.5	15.6	1800	9	AA415091	AA415091 MG0026 RC
c 21	442	15.4	1958	11	AK078953	AK078953 Mus muscu
c 22	442	15.4	3688	11	AK036443	AK036443 Mus muscu
c 23	437.5	15.3	2527	14	CD014090	CD014090 90135021
c 24	429	15.0	559	12	BM361770	BM361770 A01032-R
c 25	425	14.8	1305	14	CD014075	CD014075 90116768
c 26	425	14.8	1496	14	CD014071	CD014071 90116512
c 27	421.5	14.7	693	14	CD054007	CD054007 H002N23S
c 28	419	14.6	681	14	CD054203	CD054203 H002B22S
c 29	419	14.6	2887	11	AK077248	AK077248 Mus muscu
c 30	415	14.5	2661	11	AK033563	AK033563 Mus muscu
c 31	409	14.3	2356	11	AK037191	AK037191 Mus muscu
c 32	405	14.1	1944	11	AK078879	AK078879 Mus muscu
c 33	404	14.1	1680	29	AY414461	AY414461 Mus muscu
c 34	404	14.1	1918	14	CF111083	CF111083 Shultzomi
c 35	404	14.1	1942	11	BC015286	BC015286 Mus muscu
c 36	399	13.9	2080	11	BC001541	BC001541 Homo sapi
c 37	398	13.9	4927	11	AK040349	AK040349 Mus muscu
c 38	394	13.8	1194	14	CD014076	CD014076 90113633
c 39	394	13.8	1364	14	CD014074	CD014074 90116703
c 40	394	13.8	1585	14	CD014073	CD014073 90116676
c 41	387.5	13.5	1177	14	CD0504564	CD0504564 CDA68-G10
c 42	387.5	13.5	1680	29	AY414459	AY414459 Homo sapi
c 43	387	13.5	763	29	CG810164	CG810164 F5AAQ71TF
c 44	381	13.3	670	10	BF047018	BF047018 EST1115 M
c 45	379.5	13.3	939	13	BX854215	BX854215 BX854215

# ALIGNMENTS

RESULT 1  
LOCUS CF881106/c  
DEFINITION CF881106 trico83xn18.bl T. reesei mycelial culture, Version 6 October 2003  
ACCESSION CF881106  
VERSION CF881106.1  
KEYWORDS EST  
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)  
ORGANISM Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
REFERENCE 1 (bases 1 to 769)  
769 bp mRNA linear EST 31-OCT-2003  
trico83xn18.bl T. reesei mycelial culture, Version 6 October 2003  
CF881106  
Hypocrea jecorina CDNA clone trico83xn18, mRNA sequence.  
CF881106.1 GI:381135788

**AUTHORS** Diener S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D., Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and Dean, R.A.

**TITLE** Analysis of the protein processing and secretion pathways in a *Trichoderma reesei* EST dataset

**JOURNAL COMMENT** Unpublished (2003)  
Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seq primer: *IT-F1* primer.  
Location/Qualifiers  
source  
1. 769  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clones="trich083xn18"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

**ORIGIN**

Alignment Scores:  
Pred. No.: 7.82e-56 Length: 769  
Score: 586.00 Matches: 115  
Percent Similarity: 65.90% Conservative: 28  
Best Local Similarity: 53.00% Mismatches: 72  
Query Match: 20.46% Indels: 2  
DB: 14 Gaps: 1

US-09-943-857-4 (1-547) x CF881106 (1-769)

Qy 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrIleAlaGlyAla 125  
Db 652 CAGTGGGAAGACTGCTGACCATCAACATCACTGCTCCGAGGCGACCAACCGGGCGAC 593  
Qy 126 AsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIleGly-----SerPro 143  
Db 592 GCTCTGCGGGTCTCTTTTACATCTTCAGCGCGTTTCTCATTTGGCGGACGTGGCT 533  
Qy 144 ThrIlePheProAlaGlnMetValThrIleSerValLeuMetGlyIleHisIleIle 163  
Db 532 GCCCTGAACGACCCCTCCAGCTGCTCCAGCGGCGTCCCTTCGCGAAGCCCTTCATC 473  
Qy 164 HisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAspIleIle 183  
Db 472 TTCGCGGGCGTCAACTACCGCTCGCGGGTTCATGCTCCGAGGAGGAGATCTTC 413  
Qy 184 AlaGluGlySerGlyAsnAlaGlyLeuIleAspGlnArgLeuGlyMetGlnTrpValAla 203  
Db 412 AATGAAGGAGCGCCCAACCGCGCTTCTTGACGAGCGATGGGCTCGAATGGTGGCC 353  
Qy 204 AspAsnIleAlaGlyPheGlyGlyAspProSerIleValThrIlePheGlyGluSerAla 223  
Db 352 GATAACATTGAGCGCTTTGGCGGAACCTGCTCCAGGTGACCATCTGGGGCGAGTCCGCT 293  
Qy 224 GlySerMetSerValLeuCysHisIleIleTrpAsnAspGlyAspAsnThrIleGly 243  
Db 292 GCGTCCATTCCGCTCTGGGACCAAGCTGCTGTGTATGATGGCGATGCCACATCAACAGC 233  
Qy 244 LysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSerAspProVal 263  
Db 232 AAGCCTCTGTTTCGCGCGCCCATCATGAACCTCGGGTCCGTGACGCCCGCTCGATCGGTC 173  
Qy 264 AspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGlyCysGlySer 283

Db 172 GACTCGGAAGAGCGCGGTGTATCAGCATGTCTGTCGAGAAGGCTGATGCGACGCG 113  
Qy 284 AlaserAspIleuAlaCysLeuArgSerAlaserSerAspThrIleuLeuAspAlaThr 303  
Db 112 AGGACGACACCTGGTGTCTCCGAACTTGACCAACGACGATTCGCGGACGACG 53  
Qy 304 AsnAsnThrProGlyPheLeuAlaTyrSerSerIleuArgLeuSerTyrIleu 320  
Db 52 AATTCGTCCTCGCTGATCTGATAGTCCCTTCTCTTTCGTACCTC 2

**RESULT 2**  
CB907899/c  
LOCUS  
DEFINITION  
trich083xn18 T.reesei mycelial culture, Version 3 april Hypocrea jecorina cDNA clone trich083xn18, mRNA sequence.

**ACCESSION** CB907899  
**VERSION** CB907899.1 GI:30122557  
**KEYWORDS** ESR.  
**SOURCE** Hypocrea jecorina (anamorph: *Trichoderma reesei*)  
**ORGANISM** Hypocrea jecorina  
Bukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
**REFERENCE** 1 (bases 1 to 824)  
**AUTHORS** Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J., Kelley, A.S., Meerman, H.J., Mitchell, T., Teunissen, P.J., Ward, M., Olivarres, H.A., Teunissen, P.J., Yao, J. and Ward, M.  
**TITLE** Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*  
**JOURNAL** J Biol. Chem. 278 (34), 31988-31997 (2003)  
**MEDLINE** 22803314  
**PUBMED** 12788920  
**COMMENT** Contact: Pamela K. Foreman  
Genencor Intl.  
925 Page Mill Road, Palo Alto, CA 94304, USA  
Tel: (650) 846-7635  
Fax: (650) 621-7817  
Email: Pforeman@genencor.com  
Seq primer: *IT-F1* primer.  
Location/Qualifiers  
source  
1. 824  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clones="trich083xn18"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 3 april"  
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

**ORIGIN**

Alignment Scores:  
Pred. No.: 8.83e-56 Length: 824  
Score: 586.00 Matches: 115  
Percent Similarity: 65.90% Conservative: 28  
Best Local Similarity: 53.00% Mismatches: 72  
Query Match: 20.46% Indels: 2  
DB: 14 Gaps: 1

US-09-943-857-4 (1-547) x CB907899 (1-824)

Qy 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrIleAlaGlyAla 125  
Db 707 CAGTGGGAAGACTGCTGACCATCAACATCACTGCTCCGAGGCGACCAACCGGGCGAC 648  
Qy 126 AsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIleGly-----SerPro 143  
Db 647 GCTCTGCGGGTCTCTTTTACATCTTCAGCGGCGCTTCTCATTTGGCGGACGTGGCT 588  
Qy 144 ThrIlePheProAlaGlnMetValThrIleSerValLeuMetGlyIleHisIleIle 163

587 GCCCTGAACGACCCCTCCAGCTCGTCAGACGGCGTCCGCTTCGGCAAGCCCTTCATC 528  
 164 HisValalaValaSerTyrArgValalaSerTyrGlyPheLeuAlaGlyAspAspIleLys 183  
 527 TTCGCGCGGTCACTACCGGCTCGGGGTGGGGTTTCATGCTGGGAGGATTC 468  
 184 AlagluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTyrValala 203  
 467 AATGAAGGAGCGCAACGCGGCTTGTGACAGGCGCATGGGCTCGAATGGGTGGCC 408  
 204 AspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePheGlyGluSerAla 223  
 407 GATACATTCAGGCTTCGGGGAATCGTCCAGGTGACCATCTGGGGGACATCCGCT 348  
 224 GlySerMetSerValLeuGlyGlyAspProSerLysValThrIlePheGlyGluSerAla 243  
 347 GGCTCCATTCCTCTGGGACAGCTCGTGTGTATGATGGCATGCCATACCAAGCAC 288  
 244 LysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSerAspProVal 263  
 287 AAGCCTCTGTTCGCGCGGCATCATGAATCGGGCTTCGTCAGCGCGTGGATCGACGGC 228  
 264 AspGlyThrTyrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGlyCysGlySer 283  
 227 GACTCGGAAAGCGCAGGCGGTATCAGCATGTCTCGAAGAGCTGGATCGACGGC 168  
 284 AlaSerAspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThr 303  
 167 AGCGACGACACCTGGGTGTCTCCGAACTTACCAACAGCATTTCCGCGCAGCAGCG 108  
 304 AsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeu 320  
 107 AATTCGCTCCCTGGCTGATCTCGTACGAGTCCCTTCTTCGTACCTC 57

CF880958 719 bp mRNA linear EST 31-OCT-2003  
 trico83xc07.b1 T.reesei mycelial culture, Version 6 October 2003  
 Hypocrea jecorina cDNA clone trico83xc07, mRNA sequence.  
 CF880958  
 CF880958.1 GI:38135640  
 EST.  
 Hypocrea jecorina (anamorph: Trichoderma reesei)  
 Hypocrea jecorina  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
 1 (bases 1 to 719)  
 Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D.,  
 Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and  
 Dean, R.A.  
 Analysis of the protein processing and secretion pathways in a  
 Trichoderma reesei EST dataset  
 Unpublished (2003)  
 Contact: Ralph A. Dean  
 Fungal Genomics Laboratory  
 North Carolina State University  
 Campus Box 7251, Raleigh, NC 27695, USA  
 Tel: 919-513-0020  
 Fax: 919-513-0024  
 Email: ralph.dean@ncsu.edu  
 Seq primer: LT-Fi primer.  
 Location/Qualifiers  
 1..719  
 /organism="Hypocrea jecorina"  
 /mol\_type="mRNA"  
 /strain="QM6a"  
 /db\_xref="taxon:51453"  
 /clone="trico83xc07"  
 /dev\_stage="mycelia"  
 /clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
 /note="Vector: pREP3V; Site\_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon

## and Nitrogen sources and concentrations."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,79e-52 Length: 719  
 Score: 554.50 Matches: 117  
 Percent Similarity: 62.08% Conservative: 32  
 Best Local Similarity: 48.75% Mismatches: 88  
 Query Match: 19.36% Indels: 3  
 DB: 14 Gaps: 1

US-09-943-857-4 (1-547) x CF880958 (1-719)

QY 170 ArgValalaSerTyrGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsn 189  
 DB 1 CGGTTCGCGGCTTCGCTTCCTTCGCGGCGCGGATCTTCGCAACGCGCAGCGCAAC 60  
 QY 190 AlaGlyLeuLysAspGlnArgLeuGlyMetGlnTyrValalaAspAsnIleAlaGlyPhe 209  
 DB 61 CTCGGGCTGTGACCGACGCGCATGGGCTGCATGGGTGGCGGACAAATTCGGGCTTT 120  
 QY 210 GlyGlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeu 229  
 DB 121 GCGCGGCGACCCGACCAAGGTGACCATCTGGGGGAGTTCGCGCGCGCATTCGCTTAC 180  
 QY 230 CysHisLeuIleTyrAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAla 249  
 DB 181 GACCAGATGTGTGTACGCGGCAACGCGCATACAGCGCAAGCCCTGTTTCGCGGC 240  
 QY 250 GlyIleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsn 269  
 DB 241 GCCATCATGAATCTTCGACGCGCGTTCGCGCGGCGGCTGACTCGCCCAAGGCGCAG 300  
 QY 270 GluIleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAla 289  
 DB 301 GCCATCTACGACCAACGTCGTGAAAAAGGCGGTCAGCGCTCCGCGCGCATGCTGGCG 360  
 QY 290 CysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPhe 309  
 DB 361 TGCTTCGCGCAGGTGCGCTTACAGCAAGTTCTTCGCGCGCTCAACGTCGCGCGGCATC 420  
 QY 310 LeuAlaTyrSerSerLeu-ArgLeuSerTyrLeuProArgProAspGlyLysAsnIleT 329  
 DB 421 CTGTCTACAGTCTCTGGGGGCTGTCTACTGCGCGCGCGCGGCGGCTGCTCTCC 480  
 QY 329 hrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleLeG 349  
 DB 481 CGGACAGCGCCCGAAAGCGCTCTGTCAGCGGCGAGGTACCAAGCGCTGCGCATGATGCG 540  
 QY 349 LysAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnA 369  
 DB 541 GCGACGAGGAGACGAGGCGAGCTTCTTCATCTTCCAAAGCGACGCTCAACAGCAGCG 600  
 QY 369 laGlnAlaArgAlaTyrPheLysGln--SerPheIleHisAlaSerAspAlaGluIleA 388  
 DB 601 CGAGCTCGTGCACCTGGCTCTCGGAGTACATGTTTCCAGCGCGACGCGCGAGCTGA 660  
 QY 388 spThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThr 406  
 DB 661 CGAGCTCGTCAACACCTACGACACGCGNCAITCTCCAGCGCGCAGCCCTTTCGACG 716

RESULT 4  
 CF880958  
 LOCUS  
 DEFINITION  
 trico83xc07 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico83xc07, mRNA sequence.  
 CF880958  
 CF880958.1 GI:30122371  
 EST.  
 Hypocrea jecorina (anamorph: Trichoderma reesei)  
 Hypocrea jecorina  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

```

REFERENCE
AUTHORS
  1 (bases 1 to 774)
Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
Dunn-Coleman,N.S., Goedegebuur,P., Houfek,T.D., England,G.J.,
Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
TITLE
  Transcriptional regulation of biomass-degrading enzymes in the
  filamentous fungus Trichoderma reesei
JOURNAL
MEDLINE
PUBMED
  22803314
COMMENT
  Contact: Pamela K. Foreman
  Genencor Intl.
  925 Page Mill Road, Palo Alto, CA 94304, USA
  Tel: (650) 846-7635
  Fax: (650) 621-7817
  Email: Pforeman@genencor.com
  Seq primer: LT-F1 primer.
  Location/Qualifiers
    1..774
      /organism="Hypocrea jecorina"
      /mol_type="mRNA"
      /strain="QM6a"
      /db_xref="taxon:51453"
      /clone="trich083x07"
      /dev_stage="mycelia"
      /clone_lib="T.reesei mycelial culture, Version 3 april"
      /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
      culture grown from 24 hrs to 6 days with varying Carbon
      and Nitrogen sources and concentrations."
ORIGIN
  source
Alignment Scores:
  Pred. No.: 3,17e-52 Length: 774
  Score: 554.50 Matches: 117
  Percent Similarity: 62.08% Conservative: 32
  Best Local Similarity: 48.75% Mismatches: 88
  Query Match: 19.36% Indels: 3
  DB: 14 Gaps: 1
US-09-943-857-4 (1-547) x CB907713 (1-774)
QY 170 ArgValAlaSerTirpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsn 189
Db 56 CGCGTCGGCGGCTTCTGCTTCTGGGGCGCGGAGATCTCGCAACGGCAGCGCAAC 115
QY 190 AlaGlyLeuLysAspGlnArgLeuGlyMetGlnTirpValAlaAspAsnIleAlaGlyPhe 209
Db 116 CTCGGGCTGCTGGACCGCGCATGGGCTTGCAGTGGTGGCGGACACATCGCGCCTTT 175
QY 210 GlyGlyAspProSerIysValThrIlePheGlyGluSerAlaGlySerMetSerValLeu 229
Db 176 GCGCGCGACCCCGGACCAAGGTGACCACTCTGGGGCGAGTCCGCGGGCCCATGTCCGTCTAC 235
QY 230 CysHisLeuIleTirpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAla 249
Db 236 GACCAGATGGTGTGTACGGGGCGACCGCACGTACACGGCAAGCCCTGTTCCGGCGC 295
QY 250 GlyIleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsn 269
Db 296 GCATCATGAACCTCTGCACCGCGGTGCGCGCGAGCGGTGACTCGCCCAAGGCGCAG 355
QY 270 GluIleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAla 289
Db 356 GCATCTACGACACGTCGTGAAAAAGCGCGGTGAGCGCTCCCGGACTCGCTGGCG 415
QY 290 CysLeuArgSerAlaSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPhe 309
Db 416 TGCTTCGCGCAGGTGCGCTACAGCAAGTCTCTCGCGCGCGTCAACGTGCGCGCGGCATC 475
QY 310 LeuAlaTyrSerSerLeu--ArgLeuSerTyrLeuProArgProAspGlyLysAsnIle 329
Db 476 CTGTCTTACAGCTCCCTGGGGGGGTGTCTTACTTCTCGCGCGCGCGCGCGCTGTCTCTCC 535
QY 329 hrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleLeu 349
Db 536 CGGACACCCCGGAGCGCTCTCTGACGAGCGGTACACACGCGGTGCCCATGATTCGCCG 595
QY 349 lyaAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnA 369
Db 596 GCACACGAGGAGGAGGAGGAGCGCTCTCTCCATCTTCCAAAGCGAGCTCAACAGCAGCG 655
QY 369 laGlnAlaArgAlaTyrPheLysGln---SerPheIleHisAlaSerAspAlaGluLea 388
Db 656 CGAGCCTCGTGACCTGCTCTCGAGTACATGTTTCCAGGCGAGCGCGCGAGCTGA 715
QY 388 spThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThr 406
Db 716 CGCAGCTCGTCAACCTACGACACCGGNCATCTCCAGCGGAGCCCTTTTCGCACG 771
RESULT 5
CF885800/c 970 bp mRNA linear EST 31-OCT-2003
LOCUS
DEFINITION
  trich083xn16.b11 T.reesei mycelial culture, Version 6 October 2003
  Hypocrea jecorina cDNA clone trich083xn16, mRNA sequence.
ACCESSION
  CF885800
VERSION
  CF885800.1 GI:38140482
KEYWORDS
  EST.
SOURCE
  Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM
  Hypocrea jecorina
  Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
  Hypocreales; Hypocreales; Hypocrea; Hypocrea.
REFERENCE
  1 (bases 1 to 970)
  Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D.,
  Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and
  Dean,R.A.
  Analysis of the protein processing and secretion pathways in a
  Trichoderma reesei EST dataset
  Unpublished (2003)
  Contact: Ralph A. Dean
  Fungal Genomics Laboratory
  North Carolina State University
  Campus Box 7251, Raleigh, NC 27695, USA
  Tel: 919-513-0020
  Fax: 919-513-0024
  Email: ralph.dean@ncsu.edu
  Seq primer: LT-F1 primer.
  Location/Qualifiers
    1..970
      /organism="Hypocrea jecorina"
      /mol_type="mRNA"
      /strain="QM6a"
      /db_xref="taxon:51453"
      /clone="trich083xn16"
      /dev_stage="mycelia"
      /clone_lib="T.reesei mycelial culture, Version 6 October
      2003"
      /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
      culture grown from 24 hrs to 6 days with varying Carbon
      and Nitrogen sources and concentrations."
ORIGIN
  Alignment Scores:
  Pred. No.: 6,17e-50 Length: 970
  Score: 536.00 Matches: 120
  Percent Similarity: 55.60% Conservative: 34
  Best Local Similarity: 43.32% Mismatches: 108
  Query Match: 18.72% Indels: 16
  DB: 14 Gaps: 3
US-09-943-857-4 (1-547) x CF885800 (1-970)
QY 48 AsnLeuArgPheLysAspProValProTyrSerGlySerLeuAsnGlyGlnLysPheThr 67
Db 798 AACCTCGGTGGACACACACCAATGTGCTGTAGTTCATT----- 757
QY 68 SerTyrGlyProSerCysMetGlnGlnAsnProGluGlyThrPheGluGlnLeuGly 87

```



```

Db 756 -----GCATTCGCGCGCCCAATCATCAGTAGCACAAATTCGTCC 712
Qy 88 LysThrAlaLeuAspLeuValMetGlnSer-----LysValPheGlnAlaValLeuPro 105
Db 711 CCGCCACGCTGGCAGAGTGCATAAATCTGAGCGCTGCTTTAGCGGAGCGAGCCA 552
Qy 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125
Db 651 GT-AGTAAATGCAATATCATCAGCTTCGTAGTCCAAAGGCGCGCCACGCGGGGAC 593
Qy 126 AsnLeuProValMetLeuTrpIlePheGlyGlyPheGluIleGly-----SerPro 143
Db 592 GCTTCGCGGTCTTTTACACTCTCGAGCGCGCTTCGTCATTCGGCGGACGTCGGTT 533
Qy 144 ThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIle 163
Db 532 GCCTAAACAGCCCTCCAGCTCGTCAGACGGCGCTCGCTTCGGCAAGCCCTTCATC 473
Qy 164 HisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGlyAspIleLys 183
Db 472 TTCGCGCGGTCACTACCGCTCGCGGGTGGGGTTTCATCGCTGGGAGGAGATTC 413
Qy 184 AlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAla 203
Db 412 AATGAGGGAGCGCCAAACGCGCTTCCTTGACAGCGCATGGGGCTCGAATGGGTGGCC 353
Qy 204 AspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePheGlyGluSerAla 223
Db 352 GATAACATTGAGCGCTTTGCGGGAACTCGTCGAGGTGACCATCTGGGGCGATCCGCT 293
Qy 224 GlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsnThrTrpLysGly 243
Db 292 GGCTCATTTCCGCTCTGGGACCACTGCTGCTGATGATGGCGATGCCACATCAACGAC 233
Qy 244 LysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValProSerAspProVal 263
Db 232 AAGCTCTGTTCGCGCGCCATCATGAATCGGGCTCGGTGACGCGCCGTCGATCCGGTC 173
Qy 264 AspGlyThrTyrrGlyAsnGluIleTyrAspLeuPheValSerSerAlaGlyCysGlySer 293
Db 172 GACTCGAAAGGCAAGCGGTGATCAGCATGTCGTGAGAAAGCTGTGATCGGAGCGC 113
Qy 284 AlaSerAspLysLeuAlaCysLeuArgSerAlaSerAspThrLeuLeuAspAlaThr 303
Db 112 AGCGAGCACCCCTGGGTGTCTCCGAACTTGACCAACGACGATTTCCGCGCAGCAGCG 53
Qy 304 AsnAsnThrProGlyPheLeuAlaTyrSerSerLeuArgLeuSerTyrLeu 320
Db 52 AATTCGGTCCCTGGCTGATCTGTAAGAGTCCCTTGCTCTTTCTGTACCTC 2

RESULT 6
LOCUS AY407738
DEFINITION Mus musculus ACHE gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.
ACCESSION AY407738
VERSION AY407738.1 GI:39763709
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1723)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES
source location/Qualifiers
1..1723
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1723
/gene="ACHE"
/locus_tag="HMC2990"

Alignment Scores:
Pred. No.: 3,32e-43 Length: 1723
Score: 481.00 Matches: 162
Percent Similarity: 44.69% Conservative: 78
Best Local Similarity: 30.17% Mismatches: 177
Query Match: 16.79% Indels: 120
DB: 29 Gaps: 26

US-09-943-857-4 (1-547) x AY407738 (1-1723)
Qy 35 AlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro 54
Db 184 GCITTTCTGGCATCCCTTTGAGAGCCACCTGTGGCTCAGTAGATTATGCCACCA 243
Qy 55 Val-----ProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
Db 244 GAGCCCAAGCGCCCTGGTCAGGAGTGTGGATGCT-----ACCACCTTCCAAAT 294
Qy 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85
Db 295 GTCTGTACAGTACCTGGACCCCTGTACCTGGGTGAGGGTACTGAGATGTGAAC 354
Qy 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
Db 355 CCCAACCGAGGTG----- 369
Qy 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125
Db 370 ---AGTGAAGACTGCTGTATCTTAATGTGTGGACACCATACCCAGACCTGCTTCTCC 426
Qy 126 AsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIleGlySerProThrIle 145
Db 427 ACA---CCTGTCTCATCTGGATCTATGGGGTGGTTCCTACAGCGGAGCGGCTCTTG 483
Qy 146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIle----- 162
Db 484 -----GATGTGTATGAGCGCGCTTCTGCGCCAGGTT 516
Qy 163 -----IleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179
Db 517 GAGGGAGCTGTGTGGTATCTATGAATACCGAGTGGGAACCTTTGGCTTCTTGGCCCTA 576
Qy 180 AspAspIleLysAlaGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199
Db 577 CCAGGAACAGCA---GAAGCCCTGGCAATGTAGTCTGTGGATCAACGGGTTCCCTTG 633
Qy 200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe 219
Db 634 CAATGGGTGCAGAAATATTGCAGCCTTTGGGGCGGACCGATGTCATGACTGCTGTT 693
Qy 220 GlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239
Db 694 GGGGAGAGTGGGGTGCAGCCTCGGTGGGCATGCACATAGT----- 735

```

AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.

TITLE Direct Submission

JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.

FEATURES

source location/Qualifiers

1..1723

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

<1..>1723

/gene="ACHE"

/locus\_tag="HMC2990"

gene

ORIGIN

Alignment Scores:

Pred. No.: 3,32e-43 Length: 1723

Score: 481.00 Matches: 162

Percent Similarity: 44.69% Conservative: 78

Best Local Similarity: 30.17% Mismatches: 177

Query Match: 16.79% Indels: 120

DB: 29 Gaps: 26

US-09-943-857-4 (1-547) x AY407738 (1-1723)

Qy 35 AlaPheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro 54

Db 184 GCITTTCTGGCATCCCTTTGAGAGCCACCTGTGGCTCAGTAGATTATGCCACCA 243

Qy 55 Val-----ProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71

Db 244 GAGCCCAAGCGCCCTGGTCAGGAGTGTGGATGCT-----ACCACCTTCCAAAT 294

Qy 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85

Db 295 GTCTGTACAGTACCTGGACCCCTGTACCTGGGTGAGGGTACTGAGATGTGAAC 354

Qy 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105

Db 355 CCCAACCGAGGTG----- 369

Qy 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125

Db 370 ---AGTGAAGACTGCTGTATCTTAATGTGTGGACACCATACCCAGACCTGCTTCTCC 426

Qy 126 AsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIleGlySerProThrIle 145

Db 427 ACA---CCTGTCTCATCTGGATCTATGGGGTGGTTCCTACAGCGGAGCGGCTCTTG 483

Qy 146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIle----- 162

Db 484 -----GATGTGTATGAGCGCGCTTCTGCGCCAGGTT 516

Qy 163 -----IleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179

Db 517 GAGGGAGCTGTGTGGTATCTATGAATACCGAGTGGGAACCTTTGGCTTCTTGGCCCTA 576

Qy 180 AspAspIleLysAlaGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199

Db 577 CCAGGAACAGCA---GAAGCCCTGGCAATGTAGTCTGTGGATCAACGGGTTCCCTTG 633

Qy 200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe 219

Db 634 CAATGGGTGCAGAAATATTGCAGCCTTTGGGGCGGACCGATGTCATGACTGCTGTT 693

Qy 220 GlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239

Db 694 GGGGAGAGTGGGGTGCAGCCTCGGTGGGCATGCACATAGT----- 735

TITLE

JOURNAL Science 302 (5652), 1960-1963 (2003)

PUBMED 14671302

REFERENCE 2 (bases 1 to 1723)



```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:B230340L13"
/db_xref="MGI:2411467"
/db_xref="taxon:10090"
/clone="B230340L13"
/sex="male"
/tissue_type="corpora quadrigemina"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
108_1952

```

CDS

```

/note="unnamed protein product; acetylcholinesterase
(MGI:87876, GB|NM_009599, evidence: BLASTN, 99%,
match=1845)
putative"
/codon_start=1
/protein_id="BAC32595.1"
/db_xref="GI:26337819"
/translation="MRPPYPLHTPSLAPFLLLSLGGARAREGREDPQLVVRV
GGQLRIRLKPAGPVSAFLGIPFAEPVPGSRFPPEPRKPMWSGVLDATTFONVCYQ
YVDLYPGEGEMMNPRLSBDCLYVMTVPKPSPTPLVIWVGGEYSRAS
LDVYDGRFLAQVDEGAVLVSMNVKVGTFGLALPGSREAPCNVCLLDORALOWQENI
AAFQDPMSTVTLFGESAGASVGMHILSLPSRLFRAVLQSGTPNGPWATVSAGAR
RRATILARLVGPPGAGNDIELIAKLTPRAQDLVDHEHVLPOESIFRFSFVNV
DGFSDTEALINTGDFODLQVLGVKDEGSFLVYVPGFSKONESLISRAQLA
GVRIQVQASDLAAEAVALIYDMLHPEDTHLRDAMSAVVGDHNVCPVAQLAGRLA
ACQARVYAIFFHRASTLWLMVGMVPHGYEIEFGLPLDPSLNTVTEERLPAORLM
KYWTNPARTGDCNDPDSKSPWPVYTTAAQOVSLNKLPLEVRGLRAQTCAPNRP
LPKLLSATDTLDEAERQWAEFRWSSYVHWKQDFHYSKQERCSDL"
2158_2163
/note="putative"

```

polyA\_signal

polyA\_site

ORIGIN

```

Alignment Scores:
Pred. No.: 5e-43 Length: 2173
Score: 481.00 Matches: 162
Percent Similarity: 44.69% Conservative: 78
Best Local Similarity: 30.17% Mismatches: 177
Query Match: 16.79% Indels: 120
DB: 11 Gaps: 26

US-09-943-857-4 (1-547) x AK046080 (1-2173)

```

```

QY 35 AlapheLeuGlyIleProPheAlaGluProProValGlyAsnLeuArgPheLysAspPro 54
DB 291 GCTTTTCTGGGATCCCTTTCAGAGCCACCTGTGGGCTCAGTAGATTATGCCACCA 350
QY 55 Val-----ProTyrSerGlySerLeuAsnGlyGlnLysPheThrTyrGlyPro 71
DB 351 GAGCCCAAGCGGCCCTGGTTCAGAGTGTGGATGCT-----ACCACCTTCCAAAT 401
QY 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 35
DB 402 GTCTCTACCACTAGCTGGACACCCCTGTACCTGGGTTTGGGGTACTGAGATGTGAAC 461
QY 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
DB 462 CCCAACCGAGATTG-----476
QY 106 GlnSerGluAspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAla 125
DB 477 ---AGTGAAGACTGCTGTATCTTAATGTGTGGACACCATACCCAGACCTGCTCTCCC 533
QY 126 AsnLeuProValMetLeuThrIlePheGlyGlyGlyPheGluIleGlySerProThrIle 145
DB 534 ACA---CTGTCTCTCATCTGATCTATGCGGGTGTGTTCTACAGCGGAGCGCCCTCTTG 590
QY 146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIle-----162
DB 591 -----GATGTGTATGACGCGCCGTTTCTCGGCCCAAGTT 523

```

```

163 QY -----IleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179
624 DB GAGGGAGCTGTGGTATCTATGAACTACCGATGGGAACCTTTGGCTTCTTGGCCCTA 683
180 QY AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199
684 DB CCAGGAAGCAGA---GAAGCCCTGGCAATGTAGTCTGTGATCAACGGCTTGCCTTG 740
200 QY GlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe 219
741 DB CAATGGGTGCAAGAAATATTCAGCCTTTGGGGCGACCCGATGTCACTGCTCTCTTT 800
220 QY GlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239
801 DB GGGGAGAGTGGGGTGCAGCTCCGTGGGATGCACATCTG-----842
240 QY ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
843 DB TCCTGCCCCAGCAGGAGCCTCTTCCACAGGGCTGTCTCCAGAGTGGC-----ACACCC 896
260 QY SerAspPro-----ValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheVal 276
897 DB AATGGGCCCTGGGCCCACTGTGAGTGTGGAGAGCGCCAGCGCCACACTGCTGGCC 956
277 QY SerSerAlaGlyCys-----GlySerAlaSerAspLysLeuAlaCys 290
957 DB CGCCTTGTGGCTGTCCCCAGAGTGGCGTGGTGGCAATGACACCGAGTGTAGGCTGC 1016
291 QY LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310
1017 DB TTGAGGACAAAGCGCGCTCAGGACCTGTGGGACCACTGAGTGGCACGCTC-----CTG 1067
311 QY AlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThr 329
1068 DB CTTCAAGAAAGATCTCTCGGATTTCTTCGTGCTGTGTAGAGCGGACTTCCTCAGT 1127
330 QY AspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGly 349
1128 DB GACACACCGGAGCTCTCATCAATCTGAGATTTTCAAGACCTCAGCTGCTGTGGTGT 1187
350 QY AspGlnAsnAspGluGlyThrIlePhe-----GlyLeuSerSerLeu 363
1188 DB GTGGTGAAGGACGAGGGCTCCTACTTTCTGGTTTACGGGGTCCAGGCTTCAGCAAGAC 1247
364 QY Asn-----ValThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHis 381
1248 DB AATGAATCTCTCATCAGCGCGGCCCACTTCTCGGCTGGGGTGGGATCGGTGTACCCCAA 1307
382 QY AlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr-----ProGln 396
1308 DB GCAAGTGTACTGGCAGCGAGCTGTGTCTGTGATTTACAGACTGGCTGGCTGCACCTGAG 1367
397 QY AspIleThrGlnGlySerProPheAspThrGlyValLeu-----AsnAlaLeuThr 413
1368 DB GACCTACTCCTCAGAGATGCCATGAGTGCAGTGTAGGCGACCAACACGTTGTGTC 1427
414 QY ProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgArg 433
1428 DB CCT-----GTGGCCGAGTGGCTGGCGGACTGGCT-----1457
434 QY TyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGlnLeuSer 453
1458 DB -----GCCCAAGGGGCCCGGGTCTATGCTCTACATCTTTGGAACACGCTGCC 1502
454 QY GlyLeu-----ProIle---MetGlyThrPheHisAlaAsnAspIleValTrpGln 469
1503 DB TCCACACTGACTGTGCCCTCTGGTGGGGTGGCCCATGCTATGAATC-----1553
470 QY AspTyrLeuLeuGly-----SerGlySerValIleTyrAsnAlaPheIleAla 486
1554 DB GAGTTCATCTTTGGGCTCCCTCGATCCCTCGTGAATACACACCGAGGAGGATC 1613

```



QY	414	ProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheHleHisAlaArg	433
Db	1431	CCT-----GTGGCCAGTCGGTGGCGCATGGCT-----	1460
QY	434	TyrPheLeuAenHisPheGlnGlyThrLysTyrSerPheLeuSerLysGlnLeuSer	453
Db	1461	-----GCCCAAGGGGCCCCGGTCTATGCCTACATCTTTGAACACCGTGCC	1505
QY	454	GlyLeu-----ProIle--MetGlyThrPheHisAlaAenAspIleValTrpGln	469
Db	1506	TCCACACTGACTTGGCCCTCTGATGGGGTGCCTATGGCTATGAATC-----	1556
QY	470	AspTyrLeuLeuGly-----SerGlySerValIleTyrAsnAsnAlaPheIleAla	486
Db	1557	GAGTTTCATCTTTGGGCTCCCCTGGATCCCTCGCTGAACCTACACCACGGAGGAGGATC	1616
QY	487	PheAlaThrAspLeu-----AspProAsnThr	495
Db	1617	TTTTGCTCAGCGACTTATGAATACTGACCAATTTTCCCGCACAGGGGACCCAATGAC	1676
QY	496	Ala-----GlyLeuLeuValAsnTyrProLysTyrThrSerSerGln	510
Db	1677	CCTCGAGACTCCAAATCTCCACAGTGCCACCGTACACCACTGCCGCGCAG	1727
RESULT	9		
LOCUS	AKO42335	2184 bp mRNA linear HTC 19-SEP-2003	
DEFINITION	Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630083G07 product:acetylcholinesterase, full insert sequence.		
ACCESSION	AKO42335		
VERSION	AK042335.1 GI:26333054		
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	1	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus; Carninci,P. and Hayaashizaki,Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 95279253 10349636	
AUTHORS	2	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayaashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159	
JOURNAL MEDLINE PUBLISHED	3	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi.N., Ishii,Y., Nakamura.S., Hazama.M., Nishine.T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi.S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,B., Watahiki,W., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura.S., Kawai,J., Okazaki.Y., Muramatsu,M., Inoue,Y., Kira.A. and Hayaashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861	
REFERENCE	4		
AUTHORS	5	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)	
TITLE			
JOURNAL MEDLINE PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL MEDLINE PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL MEDLINE PUBLISHED			
REFERENCE			
AUTHORS			

of 60,770 full-length cDNAs

JOURNAL  
REFERENCE  
AUTHORS

Nature 420, 563-573 (2002)  
6 (bases 1 to 2184)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,  
Hayaehida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saigo, R., Saichou, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.

TITLE  
JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayaehida, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics  
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome  
Trust/MRC building Addenbrookes Hospital Cambridge) whose  
assistance we gratefully acknowledge.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/

FEATURES  
source

1. 2184  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:A630083G07"  
/db\_xref="MGI:2407367"  
/db\_xref="taxon:10090"  
/clone="A630083G07"  
/tissue\_type="thymus"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="3 days neonate"  
115..1959  
/note="unnamed protein product; acetylcholinesterase  
(MGI:MGI:87876, GB|NM\_009599, evidence: BLASTN, 99%,  
match=1845)  
putative"  
/codon\_start=1  
/protein\_id="BAC31228.1"  
/db\_xref="GI:26335055"  
/translation="MRPPVPLHTPSLAPFLILLILGGGARGRBDPOLLVRVR  
GGQRLKRLKAGPVSFLGIPPAEPVGSRRPPEPKRPMGSLDATTFFONVYQ  
YVDLYPFGTEGMNPNRSEDCILNVWTPYRPASTPVLIIYGGGYSGLAS  
LVYDQRFUQYEGVAVLVSMYRVGTFFGLPLPGSREAPNVGLLDQRLAQWVGENI  
AAFGDPMSTVTLFGSAGASVGMHILSPSLRPLRAVLQSGTPNGPWATVAGEAR  
REATLRLVGGCPGGAGNDTELLACLRPAQDLVDHEMLVQESIFRFSFVFW  
GDGFLSDTEALINTGDFODLVLVGVVDEGSYFLVGVPGFSKDNESLISRAQFLA  
GVRIQVQASDLAAEAVALVHTDMLHPEDPHTLRDAMSADVGDHNVCPVAOLAGRLA  
AQAARVATYIFEHRASTLTWPLMMGVPHGYEIEFGLPLDPLSNLTTERIFPAORLM  
KYWTNPARTGDPNDPDSKSPQMPFITTAQQTQVSLNKLPLVLRGLRQTCAPNRF  
LPKLASDTLDEAEQKABEFHRWSSYNVHWKQDFDHYSKQERCSDL"  
2165..2170  
/note="putative"  
2184  
/note="putative"

CDS

polya\_signal  
polyA\_site  
ORIGIN  
Alignment Scores:  
35 AlapheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspPro 54  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 357  
298 GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 357  
55 Val-----ProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 408  
358 GAGCCCAAGCGCCCTGCTGAGGAGTGTGGATGCT-----ACCACCTTCCAAAT 408  
72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 468  
409 GTCTGTACCACTACGTGGACACCTGTGGCTCACGTAGATTATGCCACCA 468  
86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 483  
469 CCCAACCGAGAGTTG-----GCTTTTCTGGGATCCCTTTGGAGCCACCTGT 483  
106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 540  
484 ---AGTGAAGACTGCTGTTATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 540  
126 AsnLeuProValMetLeuTyrPheGlyGlyGlyPheGluIleGlySerProThrIle 145  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 597  
541 ACA---CCTGTCTCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 597  
146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIle----- 162  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 630  
598 -----GATGTGTATGACGCGCTTCTGCTGCGCCAGGTT 630  
163 -----IleHisValAlaValAsnTyrArgValAlaSerTyrGlyPheLeuAlaGly 179  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 690  
631 GAGGGAGCTGTGTTGTTATCTATGAACCTACCGAGTGGGAACCTTTGGCTTCTTGGCCCTTA 690  
180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 747  
691 CCAGGAAGCAGA---GAAAGCCCTGGCAATAGTCTCTGTGGATCAACGGTTCCTGTTG 747  
200 GlnTyrValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe 219  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 807  
748 CAATGGGTGCAAGAAATATTGACGCTTTGGGGCGCACCAGCATGTGCTGACTCTGTTT 807  
220 GlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTyrAsnAspGlyAspAsn 239  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 849  
808 GGGGAGAGTGGCGGTGCGACCTCCCTGGGCGATGCACATACG----- 849  
240 ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 903  
850 TCCCTGCCAGCAGGAGCCTCTTCCAGAGGCTGTCTCCAGAGTGGC-----ACACCC 903  
260 SerAspPro-----ValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheVal 276  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 963  
904 AATGGCCCTGGGCCACTGTGTAGTGTGTGAGAGGCGCAGGCGAGGCGCCACACTGTGTCGC 963  
277 SerSerAlaGlyCys-----GlySerAlaSerAspLysLeuAlaCys 290  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 1023  
964 CGCCTTGTGGGCTGTCCCGAGGTGGCTGTGGTGGCAATGACACCGAGTGAAGCTGTC 1023  
291 LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 1074  
1024 TTGAGGACAGAGCCCGCTCAGGACCTGTGGACACGAGTGGCACCTC-----CTG 1074  
311 AlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThr 329  
|||||GCTTTTCTGGGATCCCTTTGGAGCCACCTGTGGCTCACGTAGATTATGCCACCA 1134  
1075 CCTCAAGAAAGTATCTCCGATTTTCTCGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1134

```

QY 330 AspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleGly 349
Db 1135 GACACCGGAGGCTCTCAATAGTGGAGATTTTCAAGACCTCGCGTGGTGGT 1194
QY 350 AspGlnAsnAspGluGlyThrIlePhe 363
Db 1195 GTGTGAAGACGAGGCTCTCTACTTCTGTTTACGGGTCCAGGCTTCAGCAAGAC 1254
QY 364 Asn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHis 381
Db 1255 AATGAATCTCTCATCAGCGCGGCCAGTTCCTGGCTGGGTGGCGATCGGTGTACCCCAA 1314
QY 382 AlaSerAspAlaGluIleAspThrLeuMetAlaIleTyr-----ProGln 396
Db 1315 GCAAGTCACTGGCAGCGGAGGCTGTGCTGCTGATACACAGACTGCTGCACCTTGAG 1374
QY 397 AspIleThrGlnGlySerProPheAspThrGlyValLeu-----AsnAlaLeuThr 413
Db 1375 GACCTACTCAGTACGAGATGCCATGATGCTGATGCGGACCAACGTTGTGTGC 1434
QY 414 ProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAlaArgArg 433
Db 1435 CCT-----GTGGCCAGCTGCTGGCGGACTGGCT----- 1464
QY 434 TyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLysSerLysGlnLeuSer 453
Db 1465 -----GCCCAAGGGGCCCGGCTCTATGCTTATGCTTCTTGAACACCGTGC 1509
QY 454 GlyLeu-----ProIle---MetGlyThrPheHisAlaAsnAspIleValTrpGln 469
Db 1510 TCCACACTGACTGGCCCTCTCGATGGGGTGGCCCATCGGTATGAAATC----- 1560
QY 470 AspTyrLeuLeuGly-----SerGlySerValIleTyrAsnAsnAlaPheIleAla 486
Db 1561 GAGTTCATCTTGGGTCCCTCCCTGGATCCCTGCTGACTACACCGGAGGAGGATC 1620
QY 487 PheAlaThrAspLeu-----AspProAsnThr 495
Db 1621 TTGTGTCAGGACTATGAATACTAGCAACATTTTCCCGCACAGGAGGCCCAATGAC 1680
QY 496 Ala-----GlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db 1681 CCTCGAGACTCCAAATCTCCAGTGGCCACCGTACCACTCCCGCGGAG 1731

RESULT 10
AKO50337
LOCUS
DEFINITION
Mus musculus adult male liver tumor cDNA, RIKEN full-length
enriched library, clone:C730038G20 product:butyrylcholinesterase
mRNA, full insert sequence.
ACCESSION
AKO50337
VERSION
AKO50337.1 GI:26341067
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
PUBMED
11042159
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

```

Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer Genome Res. 10 (11), 1757-1771 (2000)

20530913

11076861

4

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 3206)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akai, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully

acknowledge.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

#### FEATURES

source

1..3206  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:C730038G20"  
/db\_xref="MGI:2418059"  
/db\_xref="taxon:10090"  
/clone="C730038G20"  
/sex="male"  
/cell\_type="tumor"  
/tissue\_type="liver"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
291..2102  
/note="unnamed protein product; butyrylcholinesterase mRNA (GB|M99492, evidence: BLASTN, 99%, match=1883) putative"

CDS



```

/codon start=1
/protein_id="BAC34196.1"
/db_xref="GI:26341068"
/translation="MOTQHTKVQTHFLWILLCHPFGKSHTEEDFIITTKGRVGRG
LSPVLGVTATFLGIPYAPPPISLGRFKPQPLNKMPDINHATQVANSYQNILOAR
PGFGSEMNPNLSEDCLYLNMTIPVPPKKNATVMWITGGGFGTGTSSLPVYDGK
FLARFVIVSMNVYRGALGFAPGNPDAGNMGFLDQOLALQWVORNIAPFCGNP
KSTIIFGESAGASVSHLLCPQSYPLFTFALLESSSNAPWAKVHPPEARNRTLTLA
KFTGCKSENEMEMIKLRSKDPOEILNERNFVLPSPSILSINFGPTVDGDFLDMPHIT
LLQLGKVKQAQIILGVNKNDEGTAPLVGAPGFSKNDLSLITRAKEFQBLGNMYFCYGR
LKGKAVLYFYVDLWGESPEVYRDALDVGDNINICPALETFKKFAELNNAPFFIF
EHRSSKLPPWEMVGHYEIEFVGLGRVNYTRAEEIFSRISMTKWANFAKYGH
PRTQGNSTWMPVFTSTEQKVLTLNTEKSKYSLRAPOCQFWRLFPKPLEMTCTDID
ETQEWKAGFHRWSNMYMDNQNDYTSKESCTAL"

```

ORIGIN

Alignment Scores: 1.38e-41 Length: 3206  
 Pred. No.: 471.00 Matches: 164  
 Score: 471.00 Conserved: 74  
 Percent Similarity: 44.57%  
 Best Local Similarity: 30.71% Mismatches: 194  
 Query Match: 16.45% Indels: 102  
 DB: 11 Gaps: 24

US-09-943-857-4 (1-547) x AK050337 (1-3206)

```

Qy 22 GlyAspThrIleThrGlyLeuAsnAlaIleleAsnGluAlaPheLeuGlyIleProPhe 41
Db 441 GGTGGCAGCGTGACT-----GCTTTCTCGGTATCCCTTAT 476
Qy 42 AlaGluProGlnGlyAsnLeuArgPheLysAspProValProTyrSerGlySerLeu 61
Db 477 GCACAACTCTCTCGGTAGCTTACATGCTTCAAAAGCGCAACCTTAAACAAATGGCCT 536
Qy 62 AsnGlyGlnLysPheThrSerGlyProSerCysMetGlnGlnAsnProGluGlyThr 81
Db 537 GACATCCATAATGCCACTCAATATGCAATCTTGT-----TATCAGAACATAGCAACGCC 593
Qy 82 PheGluGluAsnLeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGln 101
Db 594 TTCCCGAGGCTTCCAGGG-----TCAGAAATGGGAAT 626
Qy 102 AlaValLeuProGlnSerGluAspCysLeuThrIleAsnValValArgPro---ProGly 120
Db 627 CCAACACAAACCTCAGTGAAGACTGCTGTATCTGATCTGATGTTGGATCCAGTACCGAAG 686
Qy 121 ThrLysAlaGlyAlaAsnLeuProValMetLeuTrpIlePheGlyGlyGlyPheGluIle 140
Db 687 CCTAAA-----AATGCCACTGCTATGATGATGATCTATGTTGGTGGCTTTCAAAC 737
Qy 141 GlySerProThrIlePheProAlaGlnMetValThrLysSerValLeuMetGlyLys 160
Db 738 GGGACCTCTTCTCTA-----CCTGTTTACGATGGGAAG 770
Qy 161 -----HisIleIleHisValAlaValAsnTyrArgValAlaSerTrp 174
Db 771 TTTCTAGTCGCTGTTCAAGAGTATTTAGTTTCGATGAATATAGGTTAGTGTCTTA 830
Qy 175 GlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAsp 194
Db 831 GGATTCCTAGCTTTTCCCGA---AATCCGATGCTCCAGGAACATGGGTTTGTGTAT 887
Qy 195 GlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSer 214
Db 888 CCAACAGTGGCACTTCAATGGGTCACAAAGAAATATAGCTCTTTTGAGGGAATCTTAA 947
Qy 215 LysValThrIlePheGlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrp 234
Db 948 AGTATACGATTTTGGAGAAAGTGCAGGGGCGAGCTTCACTAGCTTACATTTGCTC--- 1004
Qy 235 AsnAspGlyAspAsnThrTyrLysGlyLysPheProLeuPheArgAlaGlyIleMetGlnSer 254
Db 1005 -----TGCCCCCAAGAGTTATCTTTGTTTACAGAGCCATCTTCTGAAGT 1049

```

```

Qy 255 GlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsnGluIleTyrAspLeu 274
Db 1050 GGCTCC-----TCTAATGCCCTCGGAGTAAAGCATCCTCCTGAGGAAGCCAGAAACAGA 1103
Qy 275 PheValSerSerAla-----GlyCysGlySerAlaSer-----AspLysLeuAla 289
Db 1104 ACCTTCACCTTAAATTTACTGTTGCTCAAGGAAATGAGATGGAGATGATTAAA 1163
Qy 290 CysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPhe 309
Db 1164 TGCTTCGAAGTAAAGATCCTCAGGAATCTT-----CGCAATGAAGAGTTCGTCTC 1217
Qy 310 LeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThr 329
Db 1218 CCCTCTGATTCATCTTATCCATAAATTTTGGTCCAAACAGTGGATGGGATTTCTCACC 1277
Qy 330 AspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIleGly 349
Db 1278 GATATGCCCCACACACTACTCCAACTAGGAAAGTGAAGAAAGCTCAGATCTTAGTGGGA 1337
Qy 350 AspGlnAsnAspGluGlyThrIlePhe-----GlyLeuSerSerLeu 363
Db 1338 GTTAAACAAAGTGAAGGACAGCTTCTCTAGTGTACGCTGCTCGGGTTTCAGCAAGAC 1397
Qy 364 Asn-----ValThrThrAsnAlaGlnAlaArgAlaTyrPheLysGlnSerPheIleHis 381
Db 1398 AATGATAGCTTATCAACAAGGAAGAAATTTCAAGAGGTTTAAATATGATATTTCCCTGGA 1457
Qy 382 AlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGly 401
Db 1458 GTGAGCAGATTGGGCAAGGAGCAGTCTTTTCTTACTACGTGCTGTTAGGTAGCAGCAG 1517
Qy 402 SerPro-----PheAsp-----ThrGlyValLeuAsnAlaLeuThr 413
Db 1518 TCACCAAGAGTCTACCGTGACGCTTGGATGATGTTATTGGAGATTACACATCATCTGCG 1577
Qy 414 Pro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAla 431
Db 1578 CTTGCACTGGAGTTTACCAAGAAATTTGCAGAGCTTGAACAAACATGCTTTTTC----- 1631
Qy 432 ArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPheLeuSerLysGln 451
Db 1632 -----TACTTTTTCGAACAT-----CGATCTTCCAAA 1658
Qy 452 LeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyr 471
Db 1659 CTACCTTGGCCGAATGGATGGAGTGCATGCTATGATAAAT-----GAATTT 1709
Qy 472 LeuLeuGly-----SerGlySerValIleTyrAsnAsnAlaPheIleAlaPheAla 488
Db 1710 GTGTTTGGCTTACCTCTCGGAAAGAGATTAAATTATACGAGAGCTGAGGAATCTTAGT 1769
Qy 489 ThrAspLeu-----AspProAsn-----ThrAla 496
Db 1770 CGATCATTAATAAATCTGGCAAAATTTTGCAAAATATGGACATCCCAATGGGACCCAG 1829
Qy 497 GlyLeuLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db 1830 GGCAATAGCACAAATGTGGCTCTGCTTCAACAGTACTGAAACAA 1871

```

RESULT 11

AK052200 2026 bp mRNA linear HTC 20-SEP-2003  
 AK052200 LOCUS  
 DEFINITION Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched library, clone:D33006P10 product:butyrylcholinesterase mRNA, full insert sequence.  
 ACCESSION AK052200  
 VERSION AK052200.1 GI:26095058  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;





Fri Aug 6 10:53:13 2004

QY 275 pheValSerSerAla-----GlyCysGlySerAlaSer-----AspLysLeuAla 289  
Db 932 ACCTTGACCTTACCTAAATTTACTGGTGTCTCAAGGAAATAGATGGAGTATGATAA 991  
QY 290 CysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPhe 309  
Db 992 TGCCTTCGAGTAAAGATCTCAGGAATTCCT-----CGCATGGAAGGTTCGTCTC 1045  
QY 310 LeuAlaTySerSerLeuArgLeuSerTyLeuProArgProAspGlyLysAsnIleThr 329  
Db 1046 CCTCTGATCCATCTATCCATAAATTTGGTCCACAGTGGATGCGGATTTCTCACC 1105  
QY 330 AspAspMetTyLysLeuValArgAspGlyLysTyLalSerValProValIleIleGly 349  
Db 1106 GATATGCCCCACACACTACTCTCACTAGGAAAGTGAAGAAAGCTCAGATCTTAGTGGGA 1165  
QY 350 AspGlnAsnAspGluGlyThrIlePhe-----GlyLeuSerSerLeu 363  
Db 1166 GTTAAACAAAGATGAAGGACAGCTTCTCTAGTGACGTCTCGGTTTCAGCAAGAC 1225  
QY 364 Asn-----ValThrThrAsnAlaGlnAlaArgAlaTyPheLysGlnSerPheIleHis 381  
Db 1226 AATGATAGCTTATCACAGGAGGAATTTCAAGAGGTTTAAATATGATTTCTCCTGGA 1285  
QY 382 AlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyProGlnAspIleThrGlnGly 401  
Db 1286 GTGAGCAGATTGGGCAAGAGCAGTCTCTTTCTACTAGCTGGAGTGGTTAGGTGAGCAG 1345  
QY 402 SerPro-----PheAsp-----ThrGlyValLeuAsnAlaLeuThr 413  
Db 1346 TCACAGAAAGTCTACCGTGACGCTTTGGATGATGTTATGGAGATTACACATCATCTGC 1405  
QY 414 Pro-----GlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPheIleHisAla 431  
Db 1406 CTGCACTGGAGTTTACCAAGAAATTTGAGAGCTTGAACAAATGTTTTTTC----- 1459  
QY 432 ArgArgTyPheLeuAsnHisPheGlnGlyGlyThrLysTySerPheLeuSerLysGln 451  
Db 1460 -----TACTTTTTCGAACAT-----CGATCTTCCAAA 1486  
QY 452 LeuSerGlyLeuProIleMetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTy 471  
Db 1487 CTACTTGGCCGGAATGGATGGAGTGCATGCTGCTATGAATT-----GAATTT 1537  
QY 472 LeuLeuGly-----SerGlySerValIleTyAsnAsnAlaPheIleAlaPheAla 488  
Db 1538 GTGTTGGCTTACCTCTGGAGAGAGAGTAAATATATACGAGAGCTGAGGAAATCTTAGT 1597  
QY 489 ThrAspLeu-----AspProAsn-----ThrAla 496  
Db 1598 CGATCCATAATGAAAACTTGGCAAAATTTGGCAAAATATGGACATCCCATGGACCCAG 1657  
QY 497 GlyLeuValAsnTrpProLysTyThrSerSerSerGln 510  
Db 1658 GGCAATAGCAATATGGCTGCTCTTCACAGAGTACTGAAACAA 1699  
RESULT 12  
AK036259  
LOCUS 2177 bp mRNA linear HTC 19-SEP-2003  
DEFINITION Mus musculus 16 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:9630050A01 product:acetylcholinesterase, full insert sequence.  
ACCESSION AK036259  
VERSION AK036259.1 GI:26085119  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1  
REFERENCE Carninci, P. and Hayashizaki, Y.  
AUTHORS High-efficiency full-length cDNA cloning  
TITLE

Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636  
2  
REFERENCE  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
3  
REFERENCE  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861  
4  
REFERENCE  
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)  
PUBMED 11076861  
5  
REFERENCE  
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
PUBMED 12021277  
6  
REFERENCE  
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanakawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers  
1. .2177  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:9630050A01"  
/db\_xref="MGI:2400667"  
/db\_xref="taxon:10090"  
/clone="9630050A01"

/tissue\_type="cerebellum"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="16 days neonate"  
111..1955  
/note="acetylcholinesterase (MGP|MGI:87876, GB|NM\_009599,  
evidence: BLASTN, 99%, match=1845)  
putative.  
2161..2166  
/note="putative"  
2177  
/note="putative"  
ORIGIN

Alignment Scores:  
Pred. No.: 8.53e-41 Length: 2177  
Score: 461.50 Matches: 162  
Percent Similarity: 44.49% Conservative: 80  
Best Local Similarity: 29.78% Mismatches: 169  
Query Match: 16.11% Indels: 134  
DB: 11 Gaps: 26

US-09-943-857-4 (1-547) x AK036259 (1-2177)

QY 35 AlaPheLeuGlyLeuProPheAlaGluProValGlyAsnLeuArgPheLysaspPro 54  
DB 294 GCTTTCTGGGATCCCTTTCAGAGCCACCTGTGGGCTCAGTAGATTTATGCCACCA 353  
QY 55 Val-----ProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71  
DB 354 GAGCCCAAGCGCCCTGGTCAGAGGTGTGGATGCT-----ACCACCTCCAAAT 404  
QY 72 SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 35  
DB 405 GTCTGTACCACTAGCTGGACACCCCTGTACCTGGGTTTGGGTTACTGAGATGTGGAC 464  
QY 86 LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105  
DB 465 CCCAACCCGAGAGTTG----- 479  
QY 106 GlnSerGluAspCysLeuThrIleAsnValValArgProGlyThrLysAlaGlyAla 125  
DB 480 ---AGTGAAGAGTGGCTGTATCTTATGTGTGGACACCATACCCAGACCTGCTTCTCC 536  
QY 126 AsnLeuProValMetLeuThrIlePheGlyGlyPheGluIleGlySerProThrIle 145  
DB 537 ACA---CCTGTCTCATCTGGATCTATGGGTTGTTTCTACAGCGAGCGGCTCCCTG 593  
QY 146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIle----- 162  
DB 594 -----GATGTGTATGACGGCCGTTTCTGTGCCCGCAGGTT 526  
QY 163 -----IleHisValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179  
DB 627 GAGGGAGTGTGTGTATCTATGATACCTACCGAGTGGGAACCTTTGGCTTCTTGCCCTA 686  
QY 180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199  
DB 687 CCAGGAAGCAGA---GAAGCCCTGGCAATGAGTCTGTGATCAACGGCTTGCCTTG 743  
QY 200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyGlyAspProSerLysValThrIlePhe 219  
DB 744 CAATGGGTGCAGAAATAATTGACGCTTTGGGGCGGACCCGATGTCAGTCACTCTGTT 803  
QY 220 GlyGluSerAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239  
DB 804 GGGGAGAGTGGGGTGGAGCTCCGTGGGATGACATACCTG----- 345  
QY 240 ThrTrpLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259  
DB 846 TCCCTGCCCGCAGGAGGCTCTTCCACAGGGGTGCTCCACAGTGGC-----ACACCC 899  
QY 260 SerAspPro-----ValAspGlyThrTyrGlyAsnGluIleTyrAspLeuPheVal 276  
DB 940 AATGGCCCTGGGCCACTGTGAGTGTCTGAGAGGCCAGCGCAGGCGCACACTGTCTGGCC 959  
QY 277 SerSerAlaGlyCys-----GlySerAlaSerAspLysLeuAlaCys 290  
DB 960 CGCTTTGTGGCTGTCTCCCGAGTGGCGTGTGGCAATGACACCGAGCTGATAGCTGC 1019  
QY 291 LeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGlyPheLeu 310  
DB 1020 TTGAGGACAAGGCCCGCTCAGGACCTGTGGACCGACGAGTGGCAGCTC-----CTG 1070  
QY 311 AlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThr 329  
DB 1071 CCTCAAGAAAGTATCTTCGATTTTCTTCGTCCTGTGTGACGCGGACTTCCTCAGT 1130  
QY 330 AspAspMetTyrLysLeuValArgaspGlyLysTyrAla-SerValProValIleleGI 349  
DB 1131 GACACACCGAGGCTCTCATCAATCTGGAAGATTTTCAAGACCTGCAGCTGTGTGGG 1190  
QY 349 yAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsnAl 369  
DB 1191 TGTGTTGAAGGACGAGGCTCTACTTT-----CTGTTACGGGGTC 1232  
QY 369 aGlnAlaArgAlaTyrPheLysGlnSerPheIle----- 380  
DB 1233 CCAGGCTTCAGCAAGAC-AATCAATCTCTCATCAGCGCGGCCAGTTCTCGCTGGGT 1291  
QY 381 -----HisAlaSerAspAlaGluIleAspThrLeuMetAlaAlaTyr--- 394  
DB 1292 GCGGATCGGTGTACCCCAAGCAAGTGACTGGCAGCGGAGGTGTGCTCTCATTTACAC 1351  
QY 395 -----ProGlnAspIleThrGlnGlySerProPheAspThrGlyValLeu--- 409  
DB 1352 AGACTGGCTGACCTGTGAGGACCTTACTCACCTGAGAGATGCCATGAGTGCAGTGTAGG 1411  
QY 410 -----AsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAl 427  
DB 1412 CGACCACAAACGTTGTGTGCCT-----GTGGCCCGAGCTGGTGGCGACTGGC 1459  
QY 427 aPheIleHisAlaArgArgTyrPheLeuAsnHisPheGlnGlyThrLysTyrSerPh 447  
DB 1460 T-----GCCCAAGGGCGCCGGTCTATGCTTA 1486  
QY 447 eLeuSerLysGlnLeuSerGlyLeu-----ProIle---MetGlyThrPheHisAl 463  
DB 1487 CATCTTTGAACACCGTGTCTCCACTGACTGTGGCCCTCTGATGGGGGCGGCCCATAGG 1546  
QY 463 aAsnAspIleValTrpGlnAspTyrLeuLeuGly-----SerGlySerValIleTy 480  
DB 1547 CTATGAATC-----GAGTTTCATCTTTGGGCTCCCTGGATCCCTCGTGAACATA 1597  
QY 480 rAsnAsnAlaPheIleAlaPheAlaThrAspLeu----- 491  
DB 1598 CACACGAGGAGAGGATCTTTGCTCAGCGACTTATGAAATCTATGACCAATTTTGGCCG 1657  
QY 492 -----AspProAsnThrAla-----GlyLeuLeuValAsnTrpProLysTyrSer 507  
DB 1658 CACAGGGGACCCCAATGACCTCGAGACTCCAAATCTCCACAGTGGCCACCGTACACCC 1717  
QY 507 rSerSerGln 510  
DB 1718 TGCCGCGCAG 1727

RESULT 13  
AW790714  
LOCUS  
DEFINITION  
D00142-R Lambda Zap, StrataGene Blumeria graminis f. sp. hordei  
cDNA clone D00142 similar to lipase 2 precursor, mRNA sequence.  
ACCESSION  
AW790714  
VERSION  
AW790714.1 GI:13902311  
KEYWORDS  
EST.  
SOURCE  
Blumeria graminis f. sp. hordei  
Blumeria graminis f. sp. hordei  
ORGANISM  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;

QY 330 AspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleGly 349  
Db 550 GATTCTGCCACAACTTGTGAGGCTGGAAGATATGCCACGTTCTTACATTGC-GGT 608  
QY 350 AspGlnAsnAsp 353  
Db 609 GACCAAGAAGAT 620  
RESULT 14  
LOCUS CD056060 604 bp mRNA linear EST 05-JUN-2003  
DEFINITION HOI1A115 HO Hordeum vulgare cDNA clone HOI1A11 5-PRIME, mRNA  
sequence.  
ACCESSION CD056060  
VERSION CD056060.1 GI:30598520  
KEYWORDS EST.  
SOURCE Hordeum vulgare  
ORGANISM Hordeum vulgare  
REFERENCE 1 (bases 1 to 604)  
AUTHORS Zierold,U. and Schweizer,P.  
TITLE Barley ESTs from pathogen-attacked leaf epidermis  
JOURNAL Unpublished (2003)  
COMMENT Contact: Patrick Schweizer  
Transcriptome Analysis, Cyto genetics Department  
Institute of Plant Genetics and Crop Plant Research (IPK)  
Corrensstr. 3, D-06466 Gatersleben, Germany  
Tel: 0049 (0)39482-5660  
Fax: 0049 (0)39482-5595  
Email: schweiz@ipk-gatersleben.de  
Insert Length: 604 Std Error: 0.00  
Plate: 11 row: A column: 11  
Seq primer: SK.  
FEATURES  
Location/Qualifiers  
1..604  
/organism="Hordeum vulgare"  
/mol\_type="mRNA"  
/cultivar="Ingrid BC mlo-5"  
/db\_xref="GABI:707023"  
/db\_xref="taxon:4513"  
/clone="HOI1A11"  
/tissue\_type="leaf epidermis, 6 h and 24 h post  
inoculation with Blumeria graminis"  
/dev\_stage="7 d after germination"  
/lab\_host="XL10-Gold"  
/clone\_lib="HO"  
/note="Vector: pBluescript SK+; Site\_1: EcoRI (5'-end of  
cDNA); Site\_2: XhoI (3'-end of cDNA); Approximately 5 % of  
the clones correspond to cDNA from the fungi B. graminis  
hordei and tritici, respectively. Due to a cloning  
artefact caused by the kit, in most cases the EcoRI site  
is NOT present, as well as the EcoRI adapter sites upstream  
cloning. To excise the insert, restriction sites for  
EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also  
due to the cloning system used Blue/white selection for  
recombinants is not 100% reliable. Average insert size is  
1.2 kb"

Erysiphales; Erysiphaceae; Blumeria.  
1 (bases 1 to 620)  
Thomas,S.W., Rasmussen,S.W., Glaring,M.A., Rouster,J.A. and  
Oliver,R.P.  
Gene identification in the fungal pathogen Blumeria graminis by  
expressed sequence tag analysis  
Unpublished (2000)  
Contact: Rasmussen,S.W.  
Department of Yeast Genetics  
Carlsberg Laboratory  
10 Gl. Carlsbergvej, DK-2500, Copenhagen, Denmark  
Tel: 45 3327 5230  
Fax: 45 3327 4766  
Email: sw@rcr.dk  
High quality sequence stop: 620  
POLYA=NO.  
FEATURES  
Location/Qualifiers  
1..620  
/organism="Blumeria graminis f. sp. hordei"  
/mol\_type="mRNA"  
/db\_xref="taxon:62688"  
/clone="D00142"  
/cell\_type="conidia"  
/lab\_host="Hordeum vulgare"  
/clone\_lib="Lambda Zap, Stratagene"  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.06e-41 Length: 620  
Score: 461.00 Matches: 95  
Percent Similarity: 62.75% Conservatives: 33  
Best Local Similarity: 46.57% Mismatches: 76  
Query Match: 16.10% Indels: 1  
DB: 10 Gaps: 0  
US-09-943-857-4 (1-547) x AM790714 (1-620)  
QY 150 GlnMetValThrLysSerValLeuMetGlyLysHisIleLeHisValAlaValAsnTyr 169  
Db 10 AAGCTCAACAGTGGTGTGACACACGGAAGCAATTGTCTACATGGCTCACTAC 69  
QY 170 ArgValAlaSerTrpGlyPheLeuAlaGlyAspAspIleLysAlaGluGlySerGlyAsn 189  
Db 70 CGCTTGGTGGCTTGGATGTTAGCTGGAAGGAGGTGTGGATGCGGCTTACCAAC 129  
QY 190 AlaGlyLeuLysAspGlnArgLeuGlyMetGlnTrpValAlaAspAsnIleAlaGlyPhe 209  
Db 130 TTGGGACACTACGATCAGATCTTCTCTACATGGGTTCAAGATAACATCGCCAAATTC 189  
QY 210 GlyGlyAspProSerLysValThrIlePheGlyGluSerAlaGlySerMetSerValLeu 229  
Db 190 GGTGGTATCCAGAAAGGTTACTCTTTTCGGTACTCTGAGGTGTATCTCCATGG 249  
QY 230 CysHisLeuIleTrpAsnAspGlyAspAsnThrTyrLysGlyLysProLeuPheArgAla 249  
Db 250 AACCTGTTGACTGCCAAGCAGGTAACTTCGCTACACGACAAAGCTCTGTCGAGCT 309  
QY 250 GlyIleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGlyAsn 269  
Db 310 GCCATCATGGACAGTGAAGTGTATCCCACTGACCTGCCGATTTGCCAAGGCCCAA 369  
QY 270 GluIleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeuAla 289  
Db 370 AACATCTTCAACACTGCTGCTGGCTGAGGCTGTGAAGGCTGCTGACACGATTGAC 429  
QY 290 CysLeuArgSerAlaSerSerAspThrLeuLeuAlaPheAlaThrAsnAsnThrProGlyPhe 309  
Db 430 TGCCTCCGAGCTCTGCCATACGACCAATTCCTCCAAAGCTAGTACATCTGTGCTGCTC 489  
QY 310 LeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIleThr 329  
Db 490 TTCGACTACCCGCTGCTGCTCTTTCATACATTCACGATACGACGGGTATTGTGTGACA 549

```

Db      6  TGGAACTATTGACTGCTAAGGATGGTAACCTACGATACACGACAGCCCTGTGTCCGA 55
Qy      249 AlaGlyIleMetGlnSerGlyAlaMetValProSerAspProValAspGlyThrTyrGly 268
Db      66  GCTGCATATGAGCAGTGGAGTGTCTATCCACAGACCCAGCTGAGTGCCTAAAGCC 125
Qy      269 AsnGluIleTyrAspLeuPheValSerSerAlaGlyCysGlySerAlaSerAspLysLeu 288
Db      126 CAGAACATTTTACACTGCTGTGCTGGAGCTGTGAAGGCGCTGTCTGACAAAGATC 185
Qy      289 AlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsnThrProGly 308
Db      186 GACTGCCTCCAGCTGTCTACGATCAGTCTCCAGGCCAGTACATCCGTACCTGGT 245
Qy      309 PheLeuAlaTyrSerSerLeuArgLeuSerTyrLeuProArgProAspGlyLysAsnIle 328
Db      246 CTATTCGATTACCGCTGTGAGTCTCTCGTACATTCACGATATGATGGTGACCTCTTA 305
Qy      329 ThrAspAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerValProValIleIle 348
Db      306 ACCGATTCCGACATACCTCTCGGAGCGAGAAAGTATGCAACGTTCCCTACATCGTC 365
Qy      349 GlyAspGlnAsnAspGluGlyThrIlePheGlyLeuSerSerLeuAsnValThrThrAsn 368
Db      366 GGTGATCAAGAAGACGAGGTACCTCTTCAGTCTTGTAACAAGCAACATCAGACCACT 425
Qy      369 AlaGlnAlaArgAlaTyrPheLysGln---SerPheIleHisAlaSerAspAlaGluIle 387
Db      426 CAGGACATATCGACTACTTCAAGCTTTTACTCTCTCACGCTAGTATCAATTACTA 485
Qy      388 AspThrLeuMetAlaAlaTyrProGlnAspIleThrGlnGlySerProPheAspThrGly 407
Db      486 TCCGACCTTGTAACTTGTACCGAGATGACCCGCGCTGGATCGGCATACACACTGGA 545
Qy      408 ValLeuAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeu 426
Db      546 ACTCGTAACAACATATACCTCAATACAGAGACTTGGCTATTATCGTGACATG 602

RESULT 15
LOCUS   AY407736
DEFINITION Homo sapiens ACH gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.
ACCESSION AY407736
VERSION 1
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1723)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1723)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE Direct Substitution
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES
    source
        location/Qualifiers
            1..1723
                /organism="Homo sapiens"
                /mol_type="genomic DNA"

```

```

/db_xref="taxon:9606"
<1..>1723
/feature="ACH"
/locus_tag="HCM2990"

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 6,89e-40 Length: 1723
Score: 452.00 Matches: 156
Percent Similarity: 44.59% Conservative: 83
Best Local Similarity: 29.10% Mismatches: 179
Query Match: 15.78% Indels: 118
DB: 29 Gaps: 26

```

US-09-943-857-4 (1-547) x AY407736 (1-1723)

```

Qy      35  AlapheLeuGlyIleProPheAlaGluProValGlyAsnLeuArgPheLysAspPro 54
Db      184 GCITTCCTGGGCATCCCTTTGGGAGCCACCCATGGGACCCCGCTGCTTCTGCCACCG 243
Qy      55  -----ValProTyrSerGlySerLeuAsnGlyGlnLysPheThrSerTyrGlyPro 71
Db      244 GAGCCCAAGCAGCGCTTGGTCAGGGGTGGTAGACGCTACAACTTCCAGAGTGC----- 297
Qy      72  SerCysMetGln-----GlnAsnProGluGlyThrPheGluGluAsn 85
Db      298 ---TGCTACCAATATGTGGACACCCCTATACCCAGGTTTTCAGGGCACCGAGATGTGGAAC 354
Qy      86  LeuGlyLysThrAlaLeuAspLeuValMetGlnSerLysValPheGlnAlaValLeuPro 105
Db      355 CCCAACCGTGAGCTG----- 369
Qy      106 GlnSerGluAspCysLeuThrIleAsnValValArgProProGlyThrLysAlaGlyAla 125
Db      370 ---AGCGAGGACTGCTGTCCATGAATACACCGGTGGAGGCTTTGGCTTCTGCGCCCTG 423
Qy      126 AsnLeuProValMetLeuThrIlePheGlyGlyGlyPheGluIleGlySerProThrIle 145
Db      424 CCCACCCCTGCTCTGTGGATCTATGGGGTGGCTTCTACAGTGGGCGCTCTCTCTTTC 483
Qy      146 PheProAlaGlnMetValThrLysSerValLeuMetGlyLysHisIleIleHis--- 164
Db      484 -----GACGTGACGATGGCCGCTTCTGTGACAGGCC 516
Qy      165 -----ValAlaValAsnTyrArgValAlaSerTrpGlyPheLeuAlaGly 179
Db      517 GAGAGGACTGTCTGTGTCATGAATACACCGGTGGAGGCTTTGGCTTCTGCGCCCTG 576
Qy      180 AspAspIleLysAlaGluGlySerGlyAsnAlaGlyLeuLysAspGlnArgLeuGlyMet 199
Db      577 CCGGGAGCCCA---GAGGCCCGGGCAATGTGGTCTCTCTGGATCAGAGGCTGCGCCCTG 633
Qy      200 GlnTrpValAlaAspAsnIleAlaGlyPheGlyAspProSerLysValThrIlePhe 219
Db      634 CAGTGGGTGCAGGAGAGAGTGCAGCTTCGGGGTGACCCAGCATCATGACGCTGTT 693
Qy      220 GlyLysAlaGlySerMetSerValLeuCysHisLeuIleTrpAsnAspGlyAspAsn 239
Db      694 GGGGAGAGCGCGGAGCGGCTCGTGGGCATGCACCTGCTG----- 735
Qy      240 ThrTyrLysGlyLysProLeuPheArgAlaGlyIleMetGlnSerGlyAlaMetValPro 259
Db      736 TCCCGCCAGCGCGGCGCTTCCACAGGGCCGTGCTGCAGAGCGCGTGC-----CCC 789
Qy      260 SerAspProValAspGlyThrTyrGly-----AsnGluIle 271
Db      790 AATGGACCTCTG---GCCACGCTGGGATCGGAGAGCGCGCTGCGAGCGGCACGACGCTG 846
Qy      272 TyrAspLeuPheValSerSerAlaGlyCys-----GlySerAlaSer 285
Db      847 GCCACCTT-----GTGGCTGTCTCCAGGGCGGCACTGGTGGGAATGACACA 894
Qy      286 AspLysLeuAlaCysLeuArgSerAlaSerSerAspThrLeuLeuAspAlaThrAsnAsn 305

```

Fri Aug 6 10:53:13 2004

```

Db      895  GAGCTGGTAGCTGCTTCGGACAGCAGCGAGGTCCTGGTG-----AACCAC 945
Qy      306  ThrProGlyPheLeuAlaTyrSerSerLeu---ArgLeuSerTyrLeuProArgProAsp 324
Db      946  GAATGGCAGCTGCTGCTCAAGAAAGCGTCCTCCGTTCTCTCGTGGCTGTGGTAGAT 1005
Qy      325  GlyLysAsnIleThrAspMetTyrLysLeuValArgAspGlyLysTyrAlaSerVal 344
Db      1006  GGAGACTTCCTCAGTGACACCCAGAGGCCCTCATCAGCGGAGACTTCACGGCCCTG 1065
Qy      345  ProValIleIleGlyAspGlnAsnAspGluGlyThrIlePhe----- 358
Db      1066  CAGGTGCTGGTGGTGTGGTGAAGGATGAGGGCTCGTATTTCTGTTTACGGGGCCCCA 1125
Qy      359  GlyLeuSerSerLeuAsn-----ValThrAsnAlaGlnAlaAlaGlyTyrPheLys 376
Db      1126  GGCTTCAGCAAGACACAGAGTCTCTCATCAGCGGGCCGAGTTCTGCGCGGGTGCGG 1185
Qy      377  GlnSerPheIleHisAlaSerAlaGluIleAspThrLeuMetAlaAlaTyr----- 394
Db      1186  GTCGGGTTCCCCAGGTAAGTACCTGGCAGCCGAGGCTGTGTCTCTGTCATTACAGAC 1245
Qy      395  -----ProGlnAsp-----IleThrGlnGlySerProPheAspThrGlyVal 408
Db      1246  TGGCTGCATCCCGAGGACCCGCGACGCTCAGGGAGGCCCTGAGCGATGTGTGGCGAC 1305
Qy      409  LeuAsnAlaLeuThrProGlnPheLysArgIleSerAlaValLeuGlyAspLeuAlaPhe 428
Db      1306  CACATGTGTGTGCCCC-----GTGCCCCAGTGGCTGGCGGACTGGCTGCC 1353
Qy      429  IleHisAlaArg-----ArgTyrPheLeuAsnHisPheGlnGly----- 441
Db      1354  CAGGGTGGCGGGTCTACGCCCTACGCTCTTTCACACCCGCTGTTCCAGCTCTCTCGGCC 1413
Qy      442  -----GlyThrLysTyrSerPheLeuSerLysGlnLeuSerGlyLeuProIle--- 457
Db      1414  CTGTGGATGGGGGTGCCCCCAGGCTACGAGTCCAGATCGAGTTCATCTTTGGGATCCCC 1473
Qy      458  ---MetGlyThrPheHisAlaAsnAspIleValTrpGlnAspTyrLeuLeuGlySerGly 476
Db      1474  CCCTCTCGAAACTACACGGCAGAGGAGAAATCTTCGCCCGCAGCGACTGATG----- 1524
Qy      477  SerValIleTyrAsnAsnAlaPheIleAlaPheAlaThrAspLeuAspProAsnThrAla 496
Db      1525  -----CGATACTGGGCCCACTTTGCCCGCACAGGGGATCCCAATGAGCCC 1569
Qy      497  -----GlyLeuValAsnTrpProLysTyrThrSerSerSerGln 510
Db      1570  CGAGACCCCAAGGCCCCACATGGCCCCCGTACACGGCGGGGCTCAG 1617

```

Search completed: August 5, 2004, 21:18:56  
Job time : 2731 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 10:43:29 ; Search time 13 Seconds  
(without alignments)  
2190.952 Million cell updates/sec

Title: US-943-857-4

Perfect score: 2864

Sequence: 1 SMNSRGPAGRLGSVPTAKLA.....DNFTAGYDALMTNPSSFFV 547

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2779	97.0	549	1 LIP3 CANRU	P32947 candida rug
2	2513	87.7	549	1 LIP1 CANRU	P20261 candida rug
3	2431	84.9	549	1 LIP5 CANRU	P32949 candida rug
4	2366	82.6	549	1 LIP4 CANRU	P32948 candida rug
5	2314	80.8	548	1 LIP2 CANRU	P32946 candida rug
6	1092	38.1	563	1 LIP1 GEOCN	P17573 geotrichum
7	1078	37.6	563	1 LIP2 GEOCN	P22394 geotrichum
8	491.5	17.2	586	1 ACES_TORCA	P04058 torpedo cal
9	491	17.1	633	1 ACES_ELEEL	O42275 electrophor
10	489	17.1	590	1 ACES_TORMA	P07692 torpedo mar
11	482	16.8	574	1 CHLE_HORSE	P81908 equus cabal
12	481	16.8	614	1 ACES_MOUSE	P21836 mus musculu
13	476	16.6	603	1 CHLE_MOUSE	Q03311 mus musculu
14	472	16.5	629	1 ACES_LEPDE	Q27577 leptinotars
15	471	16.4	613	1 ACES_BOVIN	P23795 bos taurus
16	470	16.4	611	1 ACES_FELCA	O62763 felis silve
17	469	16.4	614	1 ACES_RAT	P37136 rattus norv
18	462	16.1	606	1 ACES_BUNFA	Q92035 bungarus fa
19	459	16.0	581	1 CHLE_RABIT	P21927 oryctolagus
20	459	16.0	602	1 CHLE_FELCA	O62760 felis silve
21	458.5	16.0	634	1 ACES_BRARE	Q98de3 brachydanio
22	458	16.0	584	1 ACES_RABIT	Q29499 oryctolagus
23	455	15.9	602	1 CHLE_PANTT	O62761 panthera t
24	453.5	15.8	737	1 ACES1 ANOAG	Q869c3 anopheles g
25	452	15.8	614	1 ACES_HUMAN	P23303 homo sapien
26	451	15.7	602	1 CHLE_HUMAN	P06276 homo sapien
27	445.5	15.6	702	1 ACES_CULPI	Q869c8 cullex pipie
28	437	15.3	620	1 ACES1_CAEER	P27459 caenorhabdi
29	431.5	15.1	489	1 PNBA_BACSU	P37967 bacillus su
30	427.5	14.9	620	1 ACES1 CAEEL	P38433 caenorhabdi
31	423	14.8	532	1 EST2_RABIT	P14943 oryctolagus
32	413	14.4	549	1 EST1_RAT	P10959 rattus norv
33	411.5	14.4	597	1 CEL_BOVIN	P30122 bos taurus

## ALIGNMENTS

### RESULT 1

ID	LIP3 CANRU	STANDARD;	PRT;	549 AA.
AC	P32947;			
DT	01-OCT-1993 (Rel. 27, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Lipase 3 precursor (EC 3.1.1.3) (Cholesterol esterase).			
GN	Lip3.			
OS	Candida rugosa (Yeast) (Candida cylindracea).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.			
OX	NCBI_TaxID=5481;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 14830;			
RX	MEDLINE=93178975; PubMed=8440480;			
RA	Lotti M., Grandori R., Fusetti F., Longhi S., Brocca S.,			
RA	Tramontano A., Alberghina L.;			
RT	"Cloning and analysis of Candida cylindracea lipase sequences."			
RL	Gene 124:45-55(1993).			
RN	[2]			
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).			
RX	MEDLINE=95308044; PubMed=7788294;			
RA	Ghosh D., Wawrzak Z., Pletnev V.Z., Li N., Kaiser R., Pangborn W.,			
RA	Joernvall H., Erman M., Duax W.L.;			
RT	"Structure of uncomplexed and linoleate-bound Candida cylindracea			
RT	cholesterol esterase.";			
RL	Structure 3:279-288(1995).			
RN	[3]			
RP	REVIEW.			
RX	MEDLINE=98451816; PubMed=9778794;			
RA	Benjamin S., Pandey A.;			
RA	"Candida rugosa lipases: molecular biology and versatility in			
RT	biotechnology.";			
RL	Yeast 14:1069-1087(1998).			
CC	- - CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a			
CC	fatty acid anion.			
CC	- - SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC				
CC	EMBL; X66006; CAA46805.1; -			
DR	PIR; JN0551; JN0551.			
DR	PDB; ICLE; 08-MAR-96.			
DR	PDB; ILLF; 07-JAN-03.			
DR	InterPro; IPR002018; CarbesteraseB.			
DR	Pfam; PF00135; Coesterase; 1.			
DR	PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.			

P56161 anopheles s  
Q64419 mesocricetu  
P16303 rattus norv  
Q64573 rattus norv  
P23953 mus musculu  
P07140 drosophila  
Q04791 anas platyr  
P36196 gallus gall  
P19835 homo sapien  
Q63010 rattus norv  
Q63108 rattus norv  
Q64285 mus musculu

34 411 14.4 664 1 ACES ANOST  
35 408.5 14.3 561 1 EST1 MESAU  
36 408 14.2 565 1 EST10 RAT  
37 402 14.0 561 1 EST4 RAT  
38 401 14.0 554 1 ESTN MOUSE  
39 401 14.0 649 1 ACES DROME  
40 396 13.8 557 1 SASB ANAPL  
41 396 13.8 767 1 ACES CHICK  
42 394 13.8 742 1 CEL HUMAN  
43 393.5 13.7 561 1 EST5 RAT  
44 392 13.7 561 1 EST3 RAT  
45 390 13.6 599 1 CEL MOUSE

DR PROSITE; PS00941; CARBOXYLESTERASE\_B\_2; 1.  
KW Cholesterol metabolism; Lipid degradation; Hydrolase; Signal;  
vw Glycoprotein; Multigene family; 3D-structure.

AN	Glycophorin, 1	75	LIPASE 3.
FT	SIGNAL	16	BY SIMILARITY.
FT	CHAIN	549	BY SIMILARITY.
FT	ACT SITE	224	BY SIMILARITY.
FT	ACT SITE	356	BY SIMILARITY.
FT	ACT SITE	464	BY SIMILARITY.
FT	ACT SITE	464	BY SIMILARITY.
FT	DISULFID	75	BY SIMILARITY.
FT	DISULFID	283	BY SIMILARITY.
FT	CARBOHYD	329	N-LINKED (GLCNAC. .)
FT	CARBOHYD	329	(POTENTIAL)
FT	CARBOHYD	365	N-LINKED (GLCNAC. .)
FT	CARBOHYD	365	(POTENTIAL)

FT	386	395
HELIX	400	402
TURN	406	407
FT	409	412
TURN	413	413
STRAND	418	429
HELIX	430	430
TURN	431	440
HELIX	446	451
STRAND	453	456
FT	458	460
TURN	463	463
STRAND	464	465
TURN	466	472
HELIX	473	473
TURN	477	478
TURN	479	482
HELIX	483	483
TURN	484	492
HELIX	495	498
STRAND	507	507
TURN	510	511
TURN	517	520
STRAND	525	528
STRAND	534	541
HELIX	544	547
HELIX	549	549
SEQUENCE	549	549
SO	SEQUENCE	SEQUENCE

Query Match 97.0%; Score 2779; DB 1; Length 549;  
Best Local Similarity 98.5%; Pred. No. 1.6e-191;  
Matches 529; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY	11	LGSVPTAKLANGDITITGLNAI	INEAF	LGI	PFAEPPVGNLRFKDP	VPVYSGSLNGOKFTSYG	70
Db	13	VAAPATKLANGDITITGLNAI	INEAF	LGI	PFAEPPVGNLRFKDP	VPVYSGSLNGOKFTSYG	72
QY	71	PSCMQQNPEGT	FEENL	GKTALD	LVMQSKVFOAVL	POSEDCLTITNVVRPPGT	KAGANLPMV 130
Db	73	PSCMQQNPEGT	FEENL	GKTALD	LVMQSKVFOAVL	POSEDCLTITNVVRPPGT	KAGANLPMV 132
QY	131	LWIPGGGFEIGSP	TIFFP	PAQWMTKSVL	MGKHI	IHVAVNRYVASWG	FLAGDDIKABSGSNA 190
Db	133	LWIPGGGFEIGSP	TIFFP	PAQWMTKSVL	MGKPI	IHVAVNRYVASWG	FLAGDDIKABSGSNA 192
QY	191	GLKDORLGMQWADNI	AGFGDP	PSKVIT	IFGBSAGS	MSVLCHLI	WNMDGNTYKGPFLFRAG 250
Db	193	GLKDORLGMQWADNI	AGFGDP	PSKVIT	IFGBSAGS	MSVLCHLI	WNMDGNTYKGPFLFRAG 252
QY	251	IMQSGAMVPSDP	VDGTG	YNEIYDL	FVSSAGCGS	ASDKLACL	RSASDITLLDATNTPGFL 310
Db	253	IMQSGAMVPSDP	VDGTG	YNEIYDL	FVSSAGCGS	ASDKLACL	RSASDITLLDATNTPGFL 312
QY	311	AYSSLRLSYLPR	PDGKNT	DDMYKLVROG	KYASVPI	IGDQND	EGTIFGLSSLNVTITNAQ 370
Db	313	AYSSLRLSYLPR	PDGKNT	DDMYKLVROG	KYASVPI	IGDQND	EGTIFGLSSLNVTITNAQ 372
QY	371	ARAYFKQSFI	HASDAE	IDTLMAAY	PODIT	IQGSPF	DTGVLNALTPOPKRISAVLGDLAFIH 430
Db	373	ARAYFKQSFI	HASDAE	IDTLMAAY	PODIT	IQGSPF	DTGVLNALTPOPKRISAVLGDLAFIH 432
QY	431	ARRYELNHF	QGGTKYS	SFLSKQL	SGLPI	MGTF	PHANDIWMQDYLLGSGSVIYNNAFIAFATD 490
Db	433	ARRYELNHF	QGGTKYS	SFLSKQL	SGLPI	MGTF	PHANDIWMQDYLLGSGSVIYNNAFIAFATD 492
QY	491	LDPNTAGLL	VNNPKY	TSSSQSG	NNLM	INALGL	YTKDKNFRTAGYDALTMTNPSFFV 547
Db	493	LDPNTAGLL	VNNPKY	TSSSQSG	NNLM	INALGL	YTKDKNFRTAGYDALTMTNPSFFV 549

RESULT 2  
LIP1 CANRU  
ID LIP1 CANRU  
AC P20261;      STANDARD;      PRT;      549 AA.



01-FEB-1991 (Rel. 17, Created)  
01-OCT-1993 (Rel. 27, Last sequence update)  
28-FEB-2003 (Rel. 41, Last annotation update)  
Lipase 1 precursor (BC 3.1.1.3).  
LIP1.  
Candida rugosa (Yeast) (Candida cylindracea).  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; mitosporic Saccharomycetales; Candida.  
NCBI\_TaxID=5481;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=ATCC 14830;  
MEDLINE=92305068; PubMed=1610906;  
Longhi S., Fusetti F., Grandori R., Lotti M., Vanoni M.,  
Alberghina L.;  
"Cloning and nucleotide sequences of two lipase genes from Candida  
cylindracea.";  
Biochim. Biophys. Acta 1131:227-232(1992).  
[2]  
SEQUENCE OF 12-549 FROM N.A., AND PARTIAL SEQUENCE.  
STRAIN=ATCC 14830 / MS-5;  
MEDLINE=89384874; PubMed=2506450;  
Kawaguchi Y., Honda H., Taniuchi-Morimura J., Iwasaki S.;  
"The codon CUG is read as serine in an asporogenic yeast Candida  
cylindracea.";  
Nature 341:164-166(1989).  
[3]  
X-RAY CRYSTALLOGRAPHY (2.06 ANGSTROMS).  
MEDLINE=93286131; PubMed=8509417;  
Grochulski P., Li Y., Schrag J.D., Bouthillier F., Smith P.,  
Harrison D., Rubin B., Cygler M.;  
"Insights into interfacial activation from an open structure of  
Candida rugosa lipase.";  
J. Biol. Chem. 268:12843-12847(1993).  
[4]  
X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS).  
MEDLINE=94190867; PubMed=8142345;  
Grochulski P., Bouthillier F., Kazlauskas R.J., Serregi A.N.,  
Schrag J.D., Zlomek E., Cygler M.;  
"Analogues of reaction intermediates identify a unique substrate  
binding site in Candida rugosa lipase.";  
Biochemistry 33:3494-3500(1994).  
[5]  
REVIEW.  
MEDLINE=98451816; PubMed=9778794;  
Benjamin S., Pandey A.;  
"Candida rugosa lipases: molecular biology and versatility in  
biotechnology.";  
Yeast 14:1069-1087(1998).  
-!- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a  
fatty acid anion.  
-!- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X64703; CAA45957.1; -;  
EMBL; X16712; CAA34684.1; -;  
PIR; S05684; S05684.  
PIR; S23448; S23448.  
PDB; 1CRL; 31-JAN-94.  
PDB; 1LPM; 20-APR-95.  
PDB; 1LPN; 20-APR-95.  
PDB; 1LPO; 20-APR-95.  
PDB; 1LPP; 20-APR-95.  
PDB; 1LPS; 08-MAR-95.  
PDB; 1TRH; 31-JAN-94.  
InterPro; IPR002018; CarboxylesteraseB.

InterPro; IPR000379; Ser esters.  
Pfam; PF00135; Coesterase; 1.  
PROSITE; PS00122; CARBOXYLESTERASE\_B\_1; 1.  
PROSITE; PS00941; CARBOXYLESTERASE\_B\_2; 1.  
KW Hydrolase; Lipid degradation; Signal; Glycoprotein; Multigene family;  
3D-structure.  
FT SIGNAL 1 15  
FT CHAIN 16 549  
FT ACT\_SITE 224 224  
FT ACT\_SITE 356 356  
FT ACT\_SITE 464 464  
FT DISULFID 75 112  
FT DISULFID 283 292  
FT CARBOHYD 329 329  
FT CARBOHYD 366 366  
FT VARIANT 398 398  
FT STRAND 18 20  
FT TURN 22 23  
FT STRAND 26 28  
FT STRAND 30 31  
FT STRAND 36 42  
FT STRAND 44 44  
FT HELIX 49 51  
FT TURN 52 53  
FT STRAND 58 58  
FT TURN 64 65  
FT STRAND 67 67  
FT STRAND 69 69  
FT STRAND 74 74  
FT TURN 80 81  
FT HELIX 88 97  
FT TURN 98 98  
FT HELIX 100 105  
FT STRAND 108 108  
FT TURN 121 122  
FT TURN 125 126  
FT STRAND 129 135  
FT TURN 139 141  
FT TURN 146 147  
FT HELIX 151 159  
FT TURN 160 161  
FT STRAND 165 169  
FT HELIX 174 178  
FT HELIX 182 187  
FT TURN 188 188  
FT TURN 190 191  
FT HELIX 192 207  
FT HELIX 208 211  
FT TURN 212 212  
FT STRAND 213 223  
FT TURN 224 224  
FT HELIX 225 235  
FT STRAND 236 239  
FT STRAND 242 243  
FT TURN 244 245  
FT STRAND 246 247  
FT STRAND 251 255  
FT TURN 265 266  
FT HELIX 268 281  
FT TURN 282 282  
FT TURN 284 285  
FT HELIX 289 295  
FT HELIX 298 306  
FT TURN 307 307  
FT TURN 311 312  
FT TURN 314 317  
FT HELIX 334 339  
FT TURN 340 341  
FT STRAND 348 353  
FT TURN 354 354  
FT STRAND 355 355  
FT TURN 356 357  
N-LINKED (GLCNAC. . .).  
G -> Q.

RESULT 3	LIPS_CANRU	STANDARD;	PRT;	549 AA.
AC	P32949;			
DT	01-OCT-1993 (Rel. 27, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	Lipase 5 precursor (EC 3.1.1.3).			
GN	LIP5.			
OS	Candida rugosa (Yeast) (Candida cylindracea).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; mitosporic Saccharomycotina; Candida.			
OX	NCBI_TaxID=5481;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=AFCC 14830;			
RC	MEDLINE=93178975; PubMed=8440490;			
RC	Lotti M., Grandori R., Fusetti F., Longhi S., Brocca S.,			
RA	Tramontano A., Alberghina L.			
RA	"Cloning and analysis of Candida cylindracea lipase sequences."			
RL	Gene 124:45-55(1993).			
RT	[2]			
RN	REVIEW.			
RP	MEDLINE=98451816; PubMed=9778794;			
RP	Benjamin S., Pandey A.;			
RT	"Candida rugosa lipases: molecular biology and versatility in			
RT	biotechnology."			
RL	Yeast 14:1069-1087(1998).			
CC	- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a			
CC	fatty acid anion.			
CC	- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
DR	EMBL; X66008; CAA46807.1; -.			
DR	PIR; JN0553; JN0553.			
DR	HSP; P32947; 1CLE.			
DR	InterPro; IPR002018; CarboxylesteraseB.			
DR	InterPro; IPR000379; Ser_gstrs.			
DR	Pfam; PF00135; Coesterase; 1.			
DR	PROSITE; PS00122; CARBOXYLESTERASE B 1; 1.			
DR	PROSITE; PS00941; CARBOXYLESTERASE B 2; 1.			
KW	Hydrolase; Lipid degradation; Signal; Glycoprotein; Multigene family.			
FT	SIGNAL 1 15			
FT	CHAIN 16 549			
FT	ACT SITE 224 224			
FT	ACT SITE 356 356			
FT	ACT SITE 464 464			
FT	DISULFID 75 112			
FT	DISULFID 283 292			
FT	DISULFID 329 329			
FT	CARBOHYD 366 366			
SQ	SEQUENCE 549 AA; 58420 MW; 0F3D04C9716F6F22 CRC64;			
Query Match	84.9%;	Score 2431;	DB 1;	Length 549;
Best Local Similarity	85.5%;	Pred. No. 1.4e-166;		
Matches 459;	Conservative 32;	Mismatches 46;	Indels 0;	Gaps 0;
QY	11	LGSVPTAKLANGDTTGLNAIINEAFGLIPFAEPVGNLRFKDPVYPSGSLNGQKFTSYG	70	
Db	13	VAAAPATTALANGDTTGLNAIINEAFGLIPFAEPVGNLRFKDPVYPSGSLNGQKFTSYG	72	
QY	71	PSCMQONPGTEENLGKLTALDLMVQSKVFQAVLPQSEDCLTINVRPDPGTAGANLPVM	130	
Db	73	PSCMQONPGTEENLPKVALDLVMQSKVFQAVLPNSEDCLTINVRPDPGTAGANLPVM	132	
QY	131	LWIFGGGFEIGSPTEIPFPAMQVTKSVLMGKHHIVAVNVRVASWGFLAGDDIKAEKSGNA	190	

```

Db 133 LWIFGGFEIGSPITFFPAQWVSKVLMGKPIIHVAVNYRLASFGLAGPDIKAEKSGNA 192
Qy 191 GLKDQRLGMQWADVNIAGFGDPKSVITIFGESAGSMSVLCHLIWNGDNTYKGLPRAG 250
Db 193 GLKDQRLGMQWADVNIAGFGDPKSVITIFGESAGSMSVLCHLLWNGDNTYKGLPRAG 252
Qy 251 IMQSGAMVPSDPVDTGTYNEIYDLFVSSAGCGSADKLAICLRASSTLLDATNNTPGFL 310
Db 253 IMQSGAMVPSDPVDTGTYNEIYDLFVSSAGCGSADKLAICLRASSTLLDATNNTPGFL 312
Qy 311 AYSSLRLSYLPRPDGKNTDDMYKLVDRGKYASVPIIGDQNDGDTIFGLSSLNVTNAQ 370
Db 313 SYSSLRLSYLPRPDGKNTDDMYKLVDRGKYASVPIIGDQNDGDTIFGLSSLNVTNAQ 372
Qy 371 ARAYFKOSFIHASDAEIDTLMAAYPODITQSGPDTGVLNALTPOFKRISAVLGDIAFIH 430
Db 373 AEAYLRKSFHATDADITALKAAVPSDVDTQSGPDTGVLNALTPOFKRISAVLGDIAFIH 432
Qy 431 ARAYFLNHFOGGTKYSLFSLQSLGSLPIMGTGTHANDIIVWQDYLLGSGSVIYNNAFIATD 490
Db 433 SRYFLNHFOGGTKYSLFSLQSLGSLPILGTGTHANDIIVWQDYLLGSGSVIYNNAFIATD 492
Qy 491 LDNTAGLLVNWPKYTSOSSQGNLMNINAGLYTGKDNFTAGYDALMTNPSFFV 547
Db 493 LDNTAGLLVNWPKYTSOSSQGNLMNINAGLYTGKDNFTAGYDALMTNPSFFV 549

```

## RESULT 4

```

ID LIP2 CANRU STANDARD; PRT; 549 AA.
AC P32946;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE Lipase 4 precursor (EC 3.1.1.3).
GN LIP4.
OS Candida rugosa (Yeast) (Candida cylindracea).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5481;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14830;
RX MEDLINE=93178975; PubMed=8440480;
RA Lotti M., Grandori R., Fusetti F., Longhi S., Brocca S.,
RA Tramontano A., Alberghina L.;
RT "Cloning and analysis of Candida cylindracea lipase sequences.";
RL Gene 124:45-55(1993).
RN [2]
RP REVIEW.
RX MEDLINE=98451816; PubMed=9778794;
RA Benjamin S., Pandey A.;
RT "Candida rugosa lipases: molecular biology and versatility in
biotechnology.";
RL Yeast 14:1069-1087(1998).
CC -!- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
fatty acid anion.
CC -!- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X66007; CAA46806.1; -
DR PIR; JN0552; JN0552.
DR HSSP; P32947; 1CLE.
DR InterPro; IPR002018; CarbesteraseB.
DR InterPro; IPR000379; Ser_estrs.

```

```

DR Pfam; PF00135; Coesterase; 1.
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
KW Hydrolase; Lipid degradation; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 15
FT CHAIN 16 549 LIPASE 4.
FT ACT_SITE 224 224 BY SIMILARITY.
FT ACT_SITE 356 356 BY SIMILARITY.
FT ACT_SITE 464 464 BY SIMILARITY.
FT DISULFID 75 112 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT CARBOHYD 366 366 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 549 AA; 58570 MW; 39C7160852F7E655 CRC64;

Query Match 82.6%; Score 2366; DB 1; Length 549;
Best Local Similarity 82.5%; Pred. No. 6,2e-162;
Matches 443; Conservative 40; Mismatches 54; Indels 0; Gaps 0;

Qy 11 LGSVPTAKLANGDTITGLNAINAEAFILGIPAEPPVGNLRFKDPVPYSGSLNGOKFTSYG 70
Db 13 VAAAPTATLANGDTITGLNAINAEAFILGIPAEPPVGNLRFKDPVPYSGSLNGOKFTSYG 72
Qy 71 PSCMQQNEGTFFBENLGKLTALDLVMSQKVPQAVLPQSEDCLTINVVPRPGTKAGANLPVM 130
Db 73 PSCMQMPLGNWDSLLPQAQMITASVLMGKPIIHVSMNYRVASWGLAGPDIAKAGSGNA 132
Qy 131 LWIFGGFEIGSPITFFPAQWVSKVLMGKPIIHVAVNYRVASWGLAGDDIIKAGSGNA 190
Db 133 VWIFGGFEVGGSSLPQAQMITASVLMGKPIIHVSMNYRVASWGLAGPDIAKAGSGNA 192
Qy 191 GLKDQRLGMQWADVNIAGFGDPKSVITIFGESAGSMSVLCHLIWNGDNTYKGLPRAG 250
Db 193 GLHDQRLGMQWADVNIAGFGDPKSVITIFGESAGSMSVLCQLWNGDNTYKGLPRAG 252
Qy 251 IMQSGAMVPSDPVDTGTYNEIYDLFVSSAGCGSADKLAICLRASSTLLDATNNTPGFL 310
Db 253 IMQSGAMVPSDPVDTGTYNEIYDLFVSSAGCGSADKLAICLRASSTLLDATNNTPGAL 312
Qy 311 AYSSLRLSYLPRPDGKNTDDMYKLVDRGKYASVPIIGDQNDGDTIFGLSSLNVTNAQ 370
Db 313 AYPSLRLSYLPRPDGDTITDDMFKLVRDGKANVPVPIIGDQNDGDTIFGLSSLNVTDAQ 372
Qy 371 ARAYFKOSFIHASDAEIDTLMAAYPODITQSGPDTGVLNALTPOFKRISAVLGDIAFIH 430
Db 373 ARQYFKESFIHASDAEIDTLMAAYPSDITQSGPDTGIFNAITPQKRIAAVLGDIAFTL 432
Qy 431 ARRYFLNHFOGGTKYSLFSLQSLGSLPIMGTGTHANDIIVWQDYLLGSGSVIYNNAFIATD 490
Db 433 PRRYFLNHFOGGTKYSLFSLQSLGSLPVGTHANDIIVWQDYLLVSHSSAVYNNAFIAFAND 492
Qy 491 LDPNTAGLLVNWPKYTSOSSQGNLMNINAGLYTGKDNFTAGYDALMTNPSFFV 547
Db 493 LDPNKAGLLVNWPKYTSOSSQGNLMNINAGLYTGKDNFTAGYDALMTNPSFFV 549

```

## RESULT 5

```

ID LIP2 CANRU STANDARD; PRT; 548 AA.
AC P32946;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE Lipase 2 precursor (EC 3.1.1.3).
GN LIP2.
OS Candida rugosa (Yeast) (Candida cylindracea).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5481;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14830;
RX MEDLINE=92305068; PubMed=1610906;
RA Longhi S., Fusetti F., Grandori R., Lotti M., Vanoni M.,

```

RA Alberghina L.;  
RT "Cloning and nucleotide sequences of two lipase genes from Candida  
RL cylindracea";  
RN Biochim. Biophys. Acta 1131:227-232(1992).  
RP [2]  
RX REVIEW.  
RA MEDLINE=98451816; PubMed=9778794;  
RT Benjamin S., Pandey A.;  
RN "Candida rugosa lipases: molecular biology and versatility in  
RT biotechnology";  
RL Yeast 14:1059-1087(1998).  
CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a  
CC fatty acid anion.  
CC -1- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@sb-sib.ch).  
CC -----  
DR EMBL; X64704; CAA45958.1; -;  
DR PIR; S32615; S32615.  
DR HSP; P32947; 1CLE.  
DR InterPro; IPR002018; CarbesteraseB.  
DR InterPro; IPR000379; Ser esters.  
DR Pfam; PF00135; Coesterase; 1.  
DR PROSITE; PS00122; CARBOXYLESTERASE B\_1; 1.  
DR PROSITE; PS00941; CARBOXYLESTERASE B\_2; 1.  
DR Hydrolase; Lipid degradation; Signal; Glycoprotein; Multigene family.  
FT SIGNAL 1 14  
FT CHAIN 15 548  
FT ACT\_SITE 223 223  
FT ACT\_SITE 355 355  
FT ACT\_SITE 463 463  
FT ACT\_SITE 74 111  
FT DISULFID 282 291  
FT DISULFID 282 365  
FT CARBOHYD 365 365  
FT SEQUENCE 548 AA; 58864 MW; E0DBCF2501E7614 CRC64;  
Query Match 80.8%; Score 2314; DB 1; Length 548;  
Best Local Similarity 80.8%; Pred. No. 3.3e-158;  
Matches 434; Conservative 35; Mismatches 68; Indels 0; Gaps 0;  
QY 11 LGSVPTAKLANGDTITGLNAINEAFIPFAEPVGNLRFKDPVPSGLNGQKFTSYG 70  
DB 12 VAAAPTATLANGDTITGLNAIVNEKFLGIPFAEPVGLRFRKPPVPSASLNGQKFTSYG 71  
QY 71 PSCMQNPFGTEENLGKATLDLMQSKVQAVLPQSECLTNVVRPPTKAGANLPVM 130  
DB 72 PSCMQNPFGSEDTLPKNAHLVLSKIFQVVLPEDECLTNVVRPPTKAGANLPVM 131  
QY 131 LMTFGGFEIGSTIPPAQMTKSVLMGKHIIHVAIVNVRVSWGFLAGDDIKAEGSGNA 190  
DB 132 LMTFGGFEIGSGSLPFGDQWAKSVLMGKPVHVSVMNVRVSWGFLAGDPIQNEGSGNA 191  
QY 191 GLKQDRLGMQVADNADNAGFGDPSKVTIFGESAGSMVLCILINVDGNTYKGPFRAG 250  
DB 192 GLHQRLLMQVADNADNAGFGDPSKVTIYESAGSMSTFVHLVNDGNDNTYKGPLFRAA 251  
QY 251 IMQSGAMVPSDVPDVTGNEIYDLFVSSACGSGASKLACLRSSDPTLADNTNTPGFL 310  
DB 252 IMQSGCMVPSDVPDVTGTYEINQVWASACGSGASKLACLRSGDPTLYQATSDTPGV 311  
QY 311 AYSLRLSYLPRPDGKNTIDMYKLVDRDGKYSVPVVIIGQNDDEGTIFGLSSLVNTTNAQ 370  
DB 312 AYPSLRLSYLPRPDGTFITDDMYALVRDGKVAHPVVIIGQNDDEGTIFGLSSLVNTTDAQ 371  
QY 371 ARAYFKQSFHASDAEIDTLMAYPQDITQSGSPDPTGVNALTPOFKRISAVILGDLAFIH 430  
DB 372 ARAYFKQSFHASDAEIDTLMAYVTSQITQSGSPDPTGTIFNAITPQFKRISALLGDLAFTL 431

QY 431 ARRYFLNHFQGTQKYSFLSKQLSGLPIMGTFFHANDIWDYLLGSGSVYNNAFIAFATD 490  
DB 432 ARRYFLNYYQGTQKYSFLSKQLSGLPVLGTFFHNDIWDYLLGSGSVYNNAFIAFAND 491  
QY 491 LDENTAGLLVNPVKYTSQSSQGNLMMINALGLYTGKDNFRTAGYDALMTNPSFFV 547  
DB 492 LDPNKAGLWNTWPTTSSQSGNNLMQINGLGLYTGKDNFRPDAYSALPNSPFFV 548  
RESULT 6  
LIP1\_GEOCN STANDARD; PRT; 563 AA.  
ID LIP1\_GEOCN  
AC P17573;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Lipase 1 precursor (EC 3.1.1.3) (GCL I).  
GN LIP1.  
OS Geotrichum candidum (Oospora lactis).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Dipodascaceae; Galactomyces.  
OX NCBI\_TaxID=27317;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=ATCC 34614;  
RC MEDLINE=90110016; PubMed=2481674;  
RX Shimada Y., Sugihara A., Tominaga Y., Iizumi T., Tsunasawa S.;  
RT "cDNA molecular cloning of Geotrichum candidum lipase";  
RL J. Biochem. 106:383-388(1989).  
RN [2]  
RP SIMILARITY TO CARBOXYLESTERASES.  
RC MEDLINE=90329888; PubMed=2115773;  
RX Slabas A.R., Windust J., Sidebottom C.M.;  
RT "Does sequence similarity of human choline esterase, Torpedo  
RT acetylcholine esterase and Geotrichum candidum lipase reveal the  
RT active site serine residue?";  
RL Biochem. J. 269:279-280(1990).  
RN [3]  
RP CHARACTERIZATION.  
RC MEDLINE=9525295; PubMed=7737187;  
RX Bertolini M.C., Schrag J.D., Cygler M., Ziomek E., Thomas D.V.,  
RA Vernet T.;  
RT "Expression and characterization of Geotrichum candidum lipase I gene.  
RT Comparison of specificity profile with lipase II";  
RL Eur. J. Biochem. 228:863-869(1995).  
CC -1- FUNCTION: Hydrolyzes all ester bonds in triglyceride and displays  
CC a high affinity for triolein. For unsaturated substrates having  
CC long fatty acyl chains (C18:2 cis-9, cis-12 and C18:3 cis-9,  
CC cis-12, cis-15) GCL I shows higher specific activity than GCL II,  
CC whereas GCL II shows higher specific activity against saturated  
CC substrates having short fatty acid chains (C8, C10, C12 and C14).  
CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a  
CC fatty acid anion.  
CC -1- SUBUNIT: Monomer.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.  
DR PIR; PNO492; ACGUCG.  
DR InterPro; IPR002018; CarbesteraseB.  
DR InterPro; IPR000379; Ser esters.  
DR Pfam; PF00135; Coesterase; 1.  
DR PROSITE; PS00122; CARBOXYLESTERASE B\_1; 1.  
DR PROSITE; PS00941; CARBOXYLESTERASE B\_2; 1.  
DR Hydrolase; Lipid degradation; Signal; Glycoprotein; Multigene family.  
KW Pyrolidone carboxylic acid.  
FT SIGNAL 1 19  
FT CHAIN 20 563  
FT MOD\_RES 20 20  
FT ACT\_SITE 236 236  
FT ACT\_SITE 373 373  
FT ACT\_SITE 482 482  
FT DISULFID 80 124  
FT DISULFID 295 307  
FT LIPASE 1.  
FT PYROLIDONE CARBOXYLIC ACID.



Fri Aug 6 10:53:13 2004

us-09-943-857-4.rsp

```

Query Match      37.6%; Score 1078; DB 1; Length 563;
Best Local Similarity 41.6%; Pred. No. 1.1e-69; Indels 42; Gaps 8;
Matches 233; Conservative 83; Mismatches 202;

QY 8 AGRGVSPTAKLANGDTITGLNALINEAFILGFPAEPVGNLRFKDPVPYSGSLNGQKFT 67
DB 15 AGVLAQAPPSLNGNEIVSGVLEKGVDTFKGIFADPPLDLRFKHPQPTGSGYQGLKAN 74

QY 68 SYGSPCMQNPETFEENLGKLTALDLMV-OSKVF-----QAVLPQSEDCLTIN 114
DB 75 DFPSPACMLQDPGNSL-----TLIDKALGLAKVPEEPRGLPYDMAKGTVMNEDCLYLN 128

QY 115 VVRPPTKAGANLPMVLMTFGGFEIGSPITRPPAQMWTKSVLMGKHIIHVAVNYVAVSW 174
DB 129 VFRPAGTKPKDALPVMWYIGGAFVYGSSAAIYFNSVYKESINQGVFVSVNYRTGPF 188

QY 175 GFLAGDDIIRAGSGNAGLKDQRLGMQWVADNIAGFGDPSKVTIPGESAGMSVTLCHLIW 234
DB 189 GFLGGDAITAGNTNAGLDQRLGKLEWSDNIANFGGDPDKVMIFGESAGAMSVAHOLIA 248

QY 235 NGDNTYKGLPFRAGIMQSGAMVPSDPVDTGTYGNEIYDLFVSSAGC---GSASDKLACL 291
DB 249 YGDDNTYNGKKLFHSAILOSGLPLPYHDSSVSGPDISYNRFAQYAGCDTSASANDTLECL 308

QY 292 RSASSDTLLDATNN-----TPGFLAYSSLRSLVLPDPDGNITDDMYKLVRDGKYA 342
DB 309 RSKSSSVLHDAQNSYDLKDLFGLLPQLFG-----PRPDGNIIPDAAYELFRSGRYA 361

QY 343 SVPIIGDQNDGDTIFGLSSLNTVTTNAQARAYFKQSFTHASDABIDITLMAAYPDITQGS 402
DB 362 KVPYISQNDGEGTAFAPVALNATTPHVKKWLQYIFVDASEASIDRVLSLPQTLSVGS 421

QY 403 PFDGVLNALTPOPKRISAVLGDIAFHARRYFLNHPGGTKYSFLSKQLSGL-PIMGTF 461
DB 422 PFRGTILNALTPOPKRVAAILSDMLFQSPRVMLSATKDVNRWYTLSTHLHNLVFFLTGF 481

QY 462 HANDIVWDYLLGSGSVLYNNAFAFAFDLPDNTAGLLVNNPKYTSSTSSGNLMMINAL 521
DB 482 HGNELIQFNVNIGPANSYLYRFLISFANHHPDENVGTNLLQWDQYT---DECKEMLEIHT 538

QY 522 GLYTGKCNFRTAGYDALMTN 541
DB 539 DNVWRTDDYRIEGISNPFETD 558

RESULT 8
ACCS TORCA STANDARD; PRT; 586 AA.
AC P04058;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Acetylcholinesterase precursor (EC 3.1.1.7) (AChE).
OS Torpedo californica (Pacific electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squala; Hynostomales; Pristigaster; Batoidae;
OC Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.
OX NCBI_TaxID=7787;
[1]
RP SEQUENCE OF 10-586 FROM N.A.
RX MEDLINE=86118676; PubMed=3753747;
RA Schumacher M., Camp S., Maulet Y., Newton M., McPhee-Quigley K.,
RA Taylor S.S., Friedmann T., Taylor P.;
RT "Primary structure of Torpedo californica acetylcholinesterase
RT deduced from its cDNA sequence."
RL Nature 319:407-409(1986).
[2]
RP SEQUENCE OF 1-9 FROM N.A.
RX MEDLINE=89066695; PubMed=3198606;
RA Schumacher M.;
RA "Multiple messenger RNA species give rise to the structural diversity
RT in acetylcholinesterase."
RL J. Biol. Chem. 263:18979-18987(1988).

```

```

[3]
RP SEQUENCE OF 552-558.
RX MEDLINE=88087239; PubMed=3335534;
RA Gibney G., MacPhee-Quigley K., Thompson B., Vedvick T., Low M.G.,
RA Taylor S.S., Taylor P.;
RT "Divergence in primary structure between the molecular forms of
RT acetylcholinesterase."
RL J. Biol. Chem. 263:1140-1145(1988).
[4]
RP ALTERNATIVE SPLICING.
RX MEDLINE=90166618; PubMed=2306366;
RA Maulet Y., Camp S., Gibney G., Rachinsky T.L., Ekstroem T.J.,
RA Taylor P.;
RT "Single gene encodes glycopospholipid-anchored and asymmetric
RT acetylcholinesterase forms: alternative coding exons contain inverted
RT repeat sequences."
RL Neuron 4:289-301(1990).
[5]
RP DISULFIDE BONDS.
RX MEDLINE=87008586; PubMed=3759980;
RA McPhee-Quigley K., Vedvick T.S., Taylor P., Taylor S.S.;
RT "Profile of the disulfide bonds in acetylcholinesterase."
RL J. Biol. Chem. 261:13565-13570(1986).
[6]
RP STRUCTURE OF THE GPI-ANCHOR.
RX MEDLINE=94079692; PubMed=8257440;
RA Mehler A., Varon L., Silman I., Homans S.W., Ferguson M.A.;
RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
RT acetylcholinesterase from the electric organ of the electric-fish,
RT Torpedo californica."
RL Biochem. J. 296:473-479(1993).
[7]
RP GPI-ANCHOR.
RX MEDLINE=96176849; PubMed=8597567;
RA Bucht G., Hjalmarsson K.;
RT "Residues in Torpedo californica acetylcholinesterase necessary for
RT processing to a glycosyl phosphatidylinositol-anchored form."
RL Biochim. Biophys. Acta 1292:223-232(1996).
[8]
RP MUTAGENESIS.
RX MEDLINE=91017542; PubMed=2217185;
RA Gibney G., Camp S., Dionne M., McPhee-Quigley K., Taylor P.;
RT "Mutagenesis of essential functional residues in
RT acetylcholinesterase."
RL Proc. Natl. Acad. Sci. U.S.A. 87:7546-7550(1990).
[9]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=91343928; PubMed=1678899;
RA Sussman J.L., Harel M., Frolow F., Oefner C., Goldman A., Tokar L.,
RA Silman I.;
RT "Atomic structure of acetylcholinesterase from Torpedo californica: a
RT prototypic acetylcholine-binding protein."
RL Science 253:872-879(1991).
[10]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF COMPLEX WITH FASCICULIN.
RX MEDLINE=96363673; PubMed=8747462;
RA Harel M., Kleywegt G.J., Ravelli R.B., Silman I., Sussman J.L.;
RT "Crystal structure of an acetylcholinesterase-fasciculin complex:
RT interaction of a three-fingered toxin from snake venom with its
RT target."
RL Structure 3:1355-1366(1995).
[11]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=97143314; PubMed=8989325;
RA Raves M.L., Harel M., Pang Y.P., Silman I., Kozikowski A.P.,
RA Sussman J.L.;
RT "Structure of acetylcholinesterase complexed with the nootropic
RT alkaloid, (-)-huperzine A."
RL Nat. Struct. Biol. 4:57-63(1997).
[12]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX MEDLINE=99249780; PubMed=10231521;
RA Bartolucci C., Perola E., Cellai L., Brufani M., Lamba D.;

```

```

RT "Back door" opening implied by the crystal structure of a
RL carbamoylated acetylcholinesterase."
RL Biochemistry 38:5714-5719(1999).
RL [13]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=99282167; PubMed=10353814;
RA Millard C.B., Kryger G., Ordentlich A., Greenblatt H.M., Harel M.,
RA Ravess M.L., Segall Y., Barak D., Shafferman A., Silman I.,
RA Sussman J.L.;
RT "Crystal structures of aged phosphonylated acetylcholinesterase:
RT nerve agent reaction products at the atomic level.";
RL Biochemistry 38:7032-7039(1999).
RL [14]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=20074924; PubMed=10606746;
RA Greenblatt H.M., Kryger G., Lewis T., Silman I., Sussman J.L.;
RA "Structure of acetylcholinesterase complexed with (-)-galanthamine at
RT 2.3-A resolution.";
RL FEBS Lett. 463:321-326(1999).
RL [15]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=99197295; PubMed=10368299;
RA Kryger G., Silman I., Sussman J.L.;
RT "Structure of acetylcholinesterase complexed with E2020 (Aricept(R)):
RT implications for the design of new anti-alzheimer drugs.";
RL Structure 7:297-307(1999).
CC -1- FUNCTION: RAPIDLY HYDROLYZES CHOLINE RELEASED INTO THE SYNAPSE.
CC MAY BE INVOLVED IN CELL-CELL INTERACTIONS.
CC -1- CATALYTIC ACTIVITY: Acetylcholine + H(2)O = choline + acetate.
CC -1- SUBUNIT: THE H FORM IS AN HOMODIMER; THE ASYMMETRIC FORM IS A
CC DISULFIDE-BONDED OLIGOMER COMPOSED OF A COLLAGENIC SUBUNIT (Q) AND
CC A VARIABLE NUMBER OF T CATALYTIC SUBUNITS.
CC -1- SUBCELLULAR LOCATION: THE H FORM IS ATTACHED TO THE MEMBRANE BY A
CC GPI-ANCHOR.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=H; Synonyms=Globular;
CC IsoId=P04058-1; Sequence=Displayed;
CC Name=T;
CC IsoId=P04058-2; Sequence=VSP_001460;
CC -1- TISSUE SPECIFICITY: ACHE IS FOUND IN THE SYNAPSES AND TO A LOWER
CC EXTENT IN EXTRAJUNCTIONAL AREAS OF MUSCLE AND NERVE, AND ON
CC BRYTHOCYTE MEMBRANES.
CC -1- PTM: An interchain disulfide bond is present in what becomes
CC position 593 of the T isoform.
CC -1- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03439; CAA27169.1; -
DR EMBL; X56516; -; NOT_ANNOTATED_CDS.
DR EMBL; X56517; -; NOT_ANNOTATED_CDS.
DR PIR; A00773; ACRYE.
DR PDB; 2ACE; 08-NOV-96.
DR PDB; 1AGU; 31-AUG-94.
DR PDB; 2ACK; 11-FEB-98.
DR PDB; 1ACL; 31-AUG-94.
DR PDB; 1AWN; 03-APR-96.
DR PDB; 1FSS; 03-SEP-97.
DR PDB; 1VOT; 16-JUN-97.
DR PDB; 1AX9; 11-FEB-98.
DR PDB; 1EVE; 22-MAR-99.
DR PDB; 1CFJ; 01-APR-99.
DR PDB; 1OCE; 18-MAY-99.
DR PDB; 2DFP; 28-JUN-99.
DR PDB; 1SOM; 25-JUN-99.

```

AC 042275;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Acetylcholinesterase precursor (EC 3.1.1.7) (ACHE).  
OS Electrophorus electricus (Electric eel).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Osteiophysi; Gymnotiformes;  
OC Electrophoridae; Electrophorus.  
OX NCBI\_TaxID=8005;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98070504; PubMed=9407087;  
RA Simon S., Massoulie J.;  
RT Cloning and expression of acetylcholinesterase from Electrophorus.  
RT Splicing pattern of the 3' exons in vivo and in transfected mammalian  
RT cells.";  
RL J. Biol. Chem. 272:33045-33055(1997).  
CC -1- FUNCTION: Rapidly hydrolyzes choline released into the synapse.  
CC -1- CATALYTIC ACTIVITY: Acetylcholine + H2O = choline + acetate.  
CC -1- SUBUNIT: DIMERS AND COLLAGEN-TAILED FORMS, IN WHICH CATALYTIC  
CC TETRAMERS ARE ASSOCIATED WITH ANCHORING PROTEINS THAT ATTACH THEM  
CC TO THE BASAL LAMINA OR TO CELL MEMBRANES. IN THE COLLAGEN-TAILED  
CC FORMS, SUBUNITS ARE ASSOCIATED WITH A SPECIFIC COLLAGEN, COLQ.  
CC WHICH TRIGGERS THE FORMATION OF ISOFORM T TETRAMERS FROM DIMERS.  
CC -1- MISCELLANEOUS: NO OTHER ISOFORMS EXIST. THIS PROTEIN CORRESPONDS  
CC TO THE T ISOFORM IN OTHER SPECIES.  
CC -1- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL: AF030422; AAB86606.1; .  
CC HSP: P04058; ISOM  
CC InterPro: IPR002018; CarbestraseB.  
CC InterPro: IPR000997; Cholinesterase.  
CC InterPro: IPR000379; Ser\_estr.  
CC Pfam: PF001135; Coesterase; 1.  
CC PRINTS: PR00878; CHOLNESTRASE.  
CC PROSITE: PS00122; CARBOXYLESTERASE\_B\_1; 1.  
CC PROSITE: PS00941; CARBOXYLESTERASE\_B\_2; 1.  
CC Hydrolase; Serine esterase; Synapse; Membrane; Nerve; Muscle; Signal;  
CC Neurotransmitter degradation; Glycoprotein.  
CC SIGNAL 1 23 POTENTIAL.  
CC CHAIN 24 633 ACETYLCHOLINESTERASE.  
CC ACT\_SITE 225 225 BY SIMILARITY.  
CC ACT\_SITE 352 352 BY SIMILARITY.  
CC ACT\_SITE 494 494 BY SIMILARITY.  
CC DISULFID 91 118 BY SIMILARITY.  
CC DISULFID 279 290 BY SIMILARITY.  
CC DISULFID 427 579 BY SIMILARITY.  
CC DISULFID 630 630 INTERCHAIN (BY SIMILARITY).  
CC CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC SEQUENCE 633 AA; 71814 MW; FC92FE7E4ADB84C3 CRC64;  
SQ  
Query Match 17.1%; Score 491; DB 1; Length 633;  
Best Local Similarity 30.3%; Pred. No. 1.5e-27;  
Matches 157; Conservative 84; Mismatches 192; Indels 86; Gaps 22;  
QY 10 RLGSVPTAKLANGTITGLNAINEARLGIIPFAEPVGNLRFKDP---VPYSGSLNGQKF 66  
DB 34 RLQGVQGTRLPVPDR-----SHVI--AFLGIIPFAEPPLGKMRFKPEPKPNVDVFDARDY 87

QY 67 TSYGFSCHQONPEGTFEENLTKALDLYMQSKVQAVLPQSEDCLTINVRPPTKAGAN 126  
DB 88 -----PSACYQYDYDTSYPGFSG-----TEMNPNRMMSDEDCLYLN-VWVPAIRPHN 133  
QY 127 LPVLMWIFGGGPEIGSPITFPDPAQVMTKSVLMKGHIH-----VAVNYRVASWAGFLAGD 180  
DB 134 LTVMWIYGGGYSSSSSL-----DVIDGRYLHSEKVVVWVSMYRVSAFGFLALN 184  
QY 181 DIKASGSGNAGLKQRLGMQWADNIAGRGDPSKVTIFGESAGSMVLCILINWDGDMT 240  
DB 185 G-SAEAPGNVGLDQRLALQWVQDNIHFPGNPKQVTIFGESAGASVGVGHLLSPD---- 239  
QY 241 YGKPLFRAGIMQSGNWPESDPVDCITYGNEIYDLFVS-----SAGC--GSASDKLACLSAS 295  
DB 240 --SRPKETRAILQSG--VPNGPWRVTVSFEARRRAIKRLGVCCPDGNDTDLIDCLRSKQ 295  
QY 296 SDTLIDATNTPGFLAYSSL-RLSYLPRPDGKNITDDMYKLVDRDGKYASVPVITIGQND 354  
DB 296 PQDLIDQEWLV---LPFSGLFRFSPVVDVGVVFPDTPPEAMLSNGNFKDTQILLGVNQNE 352  
QY 355 GTIF-----GLSSLN--VTTNAQARAYKQSPFIHASDAEIDTLMAAYPDITQGSFFDT 406  
DB 353 GSYFLIYGAPGFSKDNESLITREDFLQGVKMSVPHANEIGLEAVILQYTDMDDEDNPKN 412  
QY 407 GVNLALTPOFKRISAVLGDLAFHARRYFLNHFGQGTKYSLSKQLSGLPIMGTFRANDI 466  
DB 413 R-----EAMDDIVGDHNVVCPLOHFANMY---AQVSILOQ-----TGTASQGNL 454  
QY 467 VMQDYLLGSGSVYNNAFATATDLDENTAGLVNMPKY 505  
DB 455 GWGN--SGSASNSGNSQSVLYWFDHRASNLV--WPEW 489  
RESULT 10  
ACES TORMA STANDARD; PRT; 590 AA.  
ID ACES TORMA STANDARD; PRT; 590 AA.  
AC P07692;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Acetylcholinesterase precursor (EC 3.1.1.7) (ACHE).  
OS Torpedo marmorata (Marbled electric ray).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
OC Elasmobranchii; Squala; Hypnosquala; Priastiorajae; Batoida;  
OC Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.  
OX NCBI\_TaxID=7788;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Electric organ;  
RX MEDLINE=88004392; PubMed=2820709;  
RA Sikorav J.-L., Krejci E., Massoulie J.;  
RT "cDNA sequences of Torpedo marmorata acetylcholinesterase: primary  
RT structure of the precursor of a catalytic subunit; existence of  
RT multiple 5'-untranslated regions.";  
RL EMBO J. 6:1865-1873(1987).  
RN [2]  
RP ALTERNATIVE SPLICING.  
RX TISSUE=Electric organ;  
RX MEDLINE=89030590; PubMed=3181125;  
RA Sikorav J.-L., Duval N., Anselmet A., Bon S., Krejci E., Legay C.,  
RA Osterlund M., Reimund B., Massoulie J.;  
RT "Complex alternative splicing of acetylcholinesterase transcripts in  
RT Torpedo electric organ; primary structure of the precursor of the  
RT glycolipid-anchored dimeric form.";  
RL EMBO J. 7:2983-2993(1988).  
RN [3]  
RP SUBUNITS INTERACTIONS.  
RX TISSUE=Electric organ;  
RX MEDLINE=92371432; PubMed=1380451;  
RA Duval N., Krejci E., Grassi J., Coussen F., Massoulie J., Bon S.;  
RT "Molecular architecture of acetylcholinesterase collagen-tailed  
RT forms; construction of a glycolipid-tailed tetramer.";  
RL EMBO J. 11:3255-3261(1992).



[4]  
 RN SUBUNITS INTERACTIONS, AND REVISION TO 421.  
 RX MEDLINE=92348506; PubMed=1639848;  
 RA Duval N., Massoulie J., Bon S.;  
 RT "H and T subunits of acetylcholinesterase from Torpedo, expressed in  
 RT COS cells, generate all types of globular forms.";  
 RL J. Cell Biol. 118:641-653(1992).  
 CC -!- FUNCTION: RAPIDLY HYDROLYZES CHOLINE RELEASED INTO THE SYNAPSE.  
 CC MAY BE INVOLVED IN CELL-CELL INTERACTIONS.  
 CC -!- CATALYTIC ACTIVITY: Acetylcholine + H(2)O = choline + acetate.  
 CC -!- SUBUNIT: THE H FORM IS AN HOMODIMER; THE ASYMMETRIC FORM IS A  
 CC DISULFIDE-BONDED OLIGOMER COMPOSED OF T CATALYTIC SUBUNIT (Q) AND  
 CC A VARIABLE NUMBER OF T CATALYTIC SUBUNITS.  
 CC -!- SUBCELLULAR LOCATION: THE H FORM IS ATTACHED TO THE MEMBRANE BY A  
 CC GPI-ANCHOR.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=H; Synonyms=Globular;  
 CC IsoId=P07692-1; Sequence=Displayed;  
 CC Name=T;  
 CC IsoId=P07692-2; Sequence=VSP\_001461;  
 CC Name=3;  
 CC IsoId=P07692-3; Sequence=VSP\_001462;  
 CC -!- TISSUE SPECIFICITY: ACHE IS FOUND IN THE SYNAPSES AND TO A LOWER  
 CC EXTENT IN EXTRAJUNCTIONAL AREAS OF MUSCLE AND NERVE, AND ON  
 CC ERYTHROCYTE MEMBRANES.  
 CC -!- PTM: AN INTERCHAIN DISULFIDE BOND IS PRESENT IN WHAT BECOMES  
 CC POSITION 596 OF THE T ISOFORM (BY SIMILARITY).  
 CC -!- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [licenses@isb-sib.ch](mailto:licenses@isb-sib.ch)).  
 CC  
 DR EMBL; X05497; CAA29047.1; -;  
 DR EMBL; X13172; CAA31570.1; -;  
 DR EMBL; X13174; CAA31572.1; -;  
 DR EMBL; X13173; CAA31571.1; -;  
 DR PIR; A38868; A38868.  
 DR HSP; P04058; 1CFJ.  
 DR InterPro: IPR002018; Carboxylesterase.  
 DR InterPro: IPR000997; Cholinesterase.  
 DR InterPro: IPR000379; Ser\_estrs.  
 DR Pfam; PF00135; COesterase; 1.  
 DR PRINTS; PR00878; CHOLNESTRASE.  
 DR PROSITE; PS00122; CARBOXYLESTERASE\_B\_1; 1.  
 DR PROSITE; PS00941; CARBOXYLESTERASE\_B\_2; 1.  
 KW Hydrolase; Serine esterase; Synapse; Membrane; Nerve; Muscle; Signal;  
 KW Neurotransmitter degradation; Glycoprotein; GPI-anchor;  
 KW Alternative splicing; Lipoprotein.  
 FT SIGNAL 1 24 POTENTIAL.  
 FT CHAIN 25 567 ACETYLCHOLINESTERASE.  
 FT PROPEP 568 590 REMOVED IN MATURE FORM.  
 FT ACT\_SITE 224 224 BY SIMILARITY.  
 FT ACT\_SITE 351 351 BY SIMILARITY.  
 FT ACT\_SITE 464 464 BY SIMILARITY.  
 FT DISULFID 91 118 BY SIMILARITY.  
 FT DISULFID 278 289 BY SIMILARITY.  
 FT DISULFID 426 545 BY SIMILARITY.  
 FT DISULFID 561 561 INTERCHAIN.  
 FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT GPI-ANCHOR 567 567 GPI-anchor amidated serine.  
 FT LIPID 567 567  
 FT VARSPLIC 590  
 FT ACDELSSTGSSKGIFVFLPSILYLIFY -> ETIDEA  
 FT ERMQKTFHRSYMMHWKQDFQGRHNCAL (in  
 FT isoform T).  
 FT

FT VARSPLIC 560 590 /FTId=VSP\_001461.  
 FT ACDELSSTGSSKGIFVFLPSILYLIFY -> GNVPAF  
 FT HMKQVTPAKTVHFGVIVAHLLLSLTASDVRPLASSKWM  
 FT AHSDDLPCRRCRSHESWGRIL (in isoform 3).  
 SQ SEQUENCE 590 AA; 66744 MW; 73FAC284C9784F25 CRC64;  
 Query Match 17.1%; Score 489; DB 1; Length 590;  
 Best Local Similarity 30.8%; Pred. No. 1.8e-27;  
 Matches 161; Conservative 72; Mismatches 192; Indels 98; Gaps 23;  
 QY 35 AFIGIPFAEPVGNLRFKDP---VPYSGSLNGOKFTSYGSPCMQOQNPCTFEENLGTAL 91  
 DB 53 AFIGIPFAEPVGNMFRPEPKPWSGVNASTY---PNNQQYVDEQFPGPG--- 104  
 QY 92 DLVMQSKVFCVILPQSEDCLITNV-VRPGTKAGANLPVLMWIFGSGFELGSTIFPPAQ 150  
 DB 105 -----SEMNPNRENSEDCLYLNIWVSPRPKSAT---VMLTYGGGFYSGSSTL----- 151  
 QY 151 MVTKSVLMGKH-----IIHVAVNVVSVAGWFLA---GDDIKABGSGNAGLKQORLGMOW 202  
 DB 152 ----DVYNGKYLAYTEEVVLSVYRGAGFLALHGSQ---EAPGNMGLLDORMALQWV 204  
 QY 203 ADNIAGFGDPSPKVTIFGESAGSMVYLCHLIWMDGNTYKGFELFRAGIMQSGAMVPSDP 262  
 DB 205 HDNIQFPGDPKVTITLFGESAGRASVGMHIL-----SPGSRDLFRRAILQSGS--PNCP 256  
 QY 263 ---VDGTYGNEIYDLFVSSAGCGSADK--LACLRASASDILLDATNTPGFLAYSSL-R 316  
 DB 257 WASVSVAEGRRAVELRRNLNCLNLSNLSDELIQCLREKPKQELIDVEWNV---LPPDSIFR 313  
 QY 317 LSVLPDPCKNITDDMYKLVROGKVASVPVITGDQDEGTIFGL-----SSLNVTTN 368  
 DB 314 FEFVPVIDEFPPTSLESMLNAGNFKKITQILGVNKDEGSFFLLYCAPGSKDSKISR 373  
 QY 369 AQARAYFKOSFIHASDAEITLMAAYPQDITGSPEDTGVINALTPQFKRISAVLGDALF 428  
 DB 374 EDFMSGVGLSVPHANDLGLDAVTLQY---TDWMDNNGIKNR-----DGLDDIVGDHNV 424  
 QY 429 IHARRYFLN---HFQGGTKYSLKQLSGL---PMGTTHANDIVWDYLLSGSGSVYNN 482  
 DB 425 ICPLMHFVNKYTKFGNGITLYFFNFHRSANLVNPEWMMGVTHGYEI---EFVFGPLVLYKELN 481  
 QY 483 -----AFIAFATLDLPNTA-GLLVNWPKYTSSSQ 510  
 DB 482 YTAEBEALSRIMHYWATFATKGNPNPHSQSKWPLFTTKEQ 524  
 RESULT 11  
 CHLE\_HORSE  
 ID\_CHLE\_HORSE STANDARD; PRT; 574 AA.  
 AC P81308;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cholinesterase (EC 3.1.1.8) (Acylcholine acylhydrolase) (Choline  
 DE esterase II) (Butyrylcholine esterase) (Pseudocholinesterase) (EQ-  
 DE ECHE).  
 DE ECHE.  
 GN ECHE.  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 OX NCBI\_TaxID=9796;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Plasma;  
 RA Moorad D.R., Luo C., Garcia G.E., Doctor B.P.;  
 RT "Amino acid sequence of horse serum butyrylcholinesterase.";  
 RL (In) Doctor B.P., Taylor P., Quinn D.M., Rotundo R.L., Gentry M.K.  
 RL (eds.);  
 RL Structure and function of cholinesterases and related proteins,  
 RL pp.145-146, Plenum Press, New York and London (1998).  
 CC -!- CATALYTIC ACTIVITY: An acylcholine + H(2)O = choline + a

CC carboxylic acid anion.  
CC -i- SUBUNIT: Homotetramer. The tetramer is composed of two dimers. The  
CC two subunits in a dimer are linked by a disulfide bond.  
CC -i- TISSUE SPECIFICITY: Present in most cells except erythrocytes.  
CC -i- MISCELLANEOUS: Cholinesterase is highly reactive with  
CC organophosphate esters.  
CC  
CC -i- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.  
CC HSP; P21836; IMAA.  
CC InterPro; IPR002018; Carboxylesterase.  
CC InterPro; IPR000997; Cholinesterase.  
CC Pfam; PF00135; Coesterase; 1.  
CC PRINTS; PR00878; CHOLINESTRASE.  
CC PROSITE; PS00122; CARBOXYLESTERASE\_B\_1;  
CC PROSITE; PS00941; CARBOXYLESTERASE\_B\_2;  
CC Hydroxylase; Serine esterase; Glycoprotein.  
CC ACT SITE 198 198 BY SIMILARITY.  
CC ACT SITE 325 325 BY SIMILARITY.  
CC ACT SITE 438 438 BY SIMILARITY.  
CC DISULFID 65 92 BY SIMILARITY.  
CC DISULFID 252 263 BY SIMILARITY.  
CC DISULFID 400 519 BY SIMILARITY.  
CC DISULFID 571 571 INTERCHAIN (BY SIMILARITY).  
CC FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .).  
CC FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .).  
CC FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .).  
CC FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .).  
CC FT CARBOHYD 341 341 N-LINKED (GLCNAC. . .).  
CC FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .).  
CC FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .).  
CC FT CARBOHYD 486 486 N-LINKED (GLCNAC. . .).  
CC SEQUENCE 574 AA; 65641 MW; 07755BE9FB9CB33E CRC64;  
  
Query Match 16.8%; Score 482; DB 1; Length 574;  
Best Local Similarity 30.4%; Pred. NO. 5.6e-27;  
Matches 162; Conservative 78; Mismatches 193; Indels 100; Gaps 23;  
  
QY 22 GDTITGLNAINEAFGLIPFAEPVGNLRFDPVPSGSLGKQKFTSYGSCMOQNPRGT 81  
DB 22 GGTVT-----AFLGIPYAPQPLGRFRKFKPSQLTWSNIWATKYANSC-YQNTQS 72  
  
QY 82 FEENLGTALDVMOSKVQAVLPOSEDLTINVRP-PGYKAGANLPMVLWIFGGGPEI 140  
DB 73 PPGFLG-----SEMNPNTELSECLYLVNWIWIPAKPK- --NATWMIWYGGGQPT 120  
  
QY 141 GSPTIFPPAQMTKSVLMGK-----HIIHVAIVRVASWGLAGDDIKAGSGNAGLKD 194  
DB 121 GTSSL-----PVYDGKFLARVERVIVSMYRVGALGFLALSE-NPEAPGNMGLFD 170  
  
QY 195 QRLGQWADVADNAGFGDPSPKVTIFGESAGNSVLCHLIWNGDNTYKGLKPLFRAGIMQS 254  
DB 171 QQLALQWYOKNIARFGGNPRSVTLFGESAGAASVSLHL-----SPRSQPLFTRAILQS 224  
  
QY 255 GAMVPSDPVDGTGNGEIVDLFVSS-AGCG--SASDKLACLRSASDITLDTNNTPGFLA 311  
DB 225 GSSNAPWAVTSIYEARNRTLILAKRGGCSDNETEMIKCLRDKDPQEL---LNEVFVYP 281  
  
QY 312 YSS-LRLSVLRPDCKNITDDMYKLVRGKYASVPIIGDQNDGCTIF-----GLSLN 364  
DB 282 YDTLLSVNFGPTVDGDFLTDMPDITLQLGQKFRITQLLVGNKDEGTAFIVGAPGFSKDN 341  
  
QY 365 --VITNAQARAYFKOSFIHASDAEIDTLMAAY-----PDITQSGSPFDITGVNLALTP 414  
DB 342 NSIITRKEFQEGLKIFPPRVSEFGRESILFHYMDLWDQRAENYREALDDVVDYNIICP 401  
  
QY 415 --QFKRISAVIGDLAFIARRVFLNHFQGTGYKSLQSLGSLPMTGFHANDIYWQDYL 472  
DB 402 ALEFRKFSSELNDAFF-----YFFE-----RSTKLPMPEWMMVGMHGYEI---BVF 445  
  
QY 473 LG---SGSVIYNNA-----FIAPATLDEN-TAGLLVNNPKVTSSSQ 510  
DB 446 FGLPLERRVNYTRAEIILSRIMKRWANPAKYNFNQNTQNTSTRVFPVKSTEQ 498

RESULT 12  
ACCS MOUSE  
ID \_ACCS\_MOUSE STANDARD; PRT; 614 AA.  
AC P21836;1991 (Rel. 18, Created)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Acetylcholinesterase precursor (BC 3.1.1.7) (AChE).  
GN ACHE  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90380429; PubMed=2400605;  
RA Rachinsky T.B., Camp S., Li Y., Ekstrom T.J., Newton M., Taylor P.;  
RT "Molecular cloning of mouse acetylcholinesterase: tissue distribution  
of alternatively spliced mRNA species.";  
RL Neuron 5:317-327(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=129/SV;  
RX MEDLINE=21138439; PubMed=11239002;  
RA Wilson M.D., Riemer C., Martindale D.W., Schnupf P., Boright A.P.,  
RA Cheung T.L., Hardy D.M., Schwartz S., Scherer S.W., Tsui L.-C.,  
RA Miller W., Koop B.F.;  
RT "Comparative analysis of the gene-dense ACHE/TFP2 region on human  
chromosome 7q22 with the orthologous region on mouse chromosome 5.";  
RL Nucleic Acids Res. 29:1352-1365(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Retina;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.W., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.M., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [4]  
RP INTERACTION WITH PRIMA.  
RX MEDLINE=21664287; PubMed=11804574;  
RA Perrier A.L., Maesoulie J., Krejci E.;  
RT "PRIMA: the membrane anchor of acetylcholinesterase in the brain.";  
RL Neuron 33:275-285(2002).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS) OF COMPLEX WITH FASCICULIN.  
RX MEDLINE=96067648; PubMed=8521480;  
RA Bourne Y., Taylor P., Marchot P.;  
RT "Acetylcholinesterase inhibition by fasciculin: crystal structure of  
the complex.";  
RL Cell 83:503-512(1995).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).  
RX MEDLINE=99115643; PubMed=9915834;  
RA Bourne Y., Taylor P., Bougis P.E., Marchot P.;  
RT "Crystal structure of mouse acetylcholinesterase. A peripheral site-

occluding loop in a tetrameric assembly.";

RL J. Biol. Chem. 274:2963-2970(1999).

CC -!- FUNCTION: Rapidly hydrolyzes choline released into the synapse.

CC -!- CATALYTIC ACTIVITY: Acetylcholine + H<sub>2</sub>O = choline + acetate.

CC -!- SUBUNIT: Isoform H generates GPI-anchored dimers; disulfide

CC linked. Isoform T generates multiple structures, ranging from

CC monomers and dimers to collagen-tailed and hydrophobic-tailed

CC forms, in which catalytic tetramers are associated with anchoring

CC proteins that attach them to the basal lamina or to cell

CC membranes. In the collagen-tailed forms, isoform T subunits are

CC associated with a specific collagen, COLQ, which triggers the

CC formation of isoform T tetramers, from monomers and dimers (by

CC similarity). Interacts with PRIMA1. The interaction with PRIMA1 is

CC required to anchor it to the basal lamina of cells and organize

CC into tetramers.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=T;

CC IsoId=P21836-1; Sequences=Displayed;

CC Name=H;

CC IsoId=P21836-2; Sequence=Not described;

CC Note=No experimental confirmation available;

CC -!- TISSUE SPECIFICITY: Predominates in most expressing tissues except

CC erythrocytes where a glycopospholipid-attached form of AChE

CC predominates.

CC -!- MISCELLANEOUS: Synapses usually contain asymmetric molecules of

CC cholinesterase, with a collagen-like part disulfide-bonded to the

CC catalytic part. A different, globular type of cholinesterase

CC occurs on the outer surfaces of cell membranes, including those of

CC erythrocytes.

CC -!- MISCELLANEOUS: This is the catalytic subunit of an asymmetric or

CC soluble form of AChE.

CC -!- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X56518; CAA39867.1; -;

DR EMBL; AF312033; AAK28816.1; -;

DR EMBL; BC046327; AAH46327.1; -;

DR PIR; JH0314; JH0314

DR PDB; 1MAH; 03-APR-96.

DR PDB; 1MAA; 20-APR-99.

DR PDB; 1C2B; 29-DEC-99.

DR PDB; 1C2O; 19-JAN-00.

DR PDB; 1J06; 04-FEB-03.

DR PDB; 1J07; 04-FEB-03.

DR PDB; 1NSM; 04-FEB-03.

DR PDB; 1NSR; 04-FEB-03.

DR MGD; MGI:87876; Ache.

DR GO; GO:0045202; C:synaptic junction; IDA.

DR InterPro; IPR002018; CarboxylesteraseB.

DR InterPro; IPR000997; Cholinesterase.

DR Pfam; PF00135; Coesterase; 1.

DR PRINTS; PR00878; CHOLINESTRASE.

DR PROSITE; PS00122; CARBOXYLESTERASE B\_1; 1.

DR PROSITE; PS00941; CARBOXYLESTERASE B\_2; 1.

KW Hydrolase; Serine esterase; Synapse; Membrane; Nerve; Muscle; Signal;

KW Neurotransmitter degradation; Glycoprotein; Alternative splicing;

KW 3D-structure.

FT SIGNAL 1 31

FT CHAIN 32 614 ACETYLCHOLINESTERASE.

FT ACT\_SITE 234 234

FT ACT\_SITE 365 365

FT ACT\_SITE 478 478

FT DISULFID 100 127

FT DISULFID 288 303

611 611 INTERCHAIN (BY SIMILARITY).

296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).

381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).

495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).

39 39

43 43

44 44

45 45

49 49

55 55

57 57

67 67

69 69

76 76

77 77

83 83

92 92

94 94

100 100

109 109

110 110

115 115

116 116

117 117

124 124

135 135

143 143

153 153

155 155

160 160

162 162

164 164

173 173

180 180

185 185

194 194

202 202

217 217

221 221

222 222

233 233

234 234

244 244

250 250

252 252

255 255

270 270

285 285

287 287

306 306

315 315

319 319

333 333

348 348

354 354

362 362

364 364

370 370

371 371

375 375

387 387

399 399

440 440

611 611 DISULFID

296 296 FT CARBOHYD

381 381 FT CARBOHYD

495 495 FT CARBOHYD

39 39 FT HELIX

43 43 FT STRAND

44 44 FT TURN

45 44 FT STRAND

49 46 FT STRAND

55 51 FT STRAND

57 56 FT TURN

67 58 FT STRAND

69 69 FT STRAND

76 74 FT HELIX

77 77 FT TURN

83 83 FT STRAND

92 90 FT STRAND

94 94 FT STRAND

100 99 FT STRAND

109 109 FT TURN

110 109 FT TURN

115 112 FT HELIX

116 116 FT TURN

117 117 FT STRAND

124 123 FT STRAND

135 129 FT STRAND

143 143 FT STRAND

153 153 FT TURN

155 155 FT TURN

160 159 FT TURN

162 162 FT HELIX

164 166 FT HELIX

173 176 FT STRAND

180 176 FT STRAND

185 185 FT HELIX

194 193 FT TURN

202 202 FT HELIX

217 217 FT HELIX

221 221 FT TURN

222 222 FT STRAND

233 223 FT TURN

234 234 FT HELIX

244 235 FT HELIX

250 247 FT TURN

252 251 FT STRAND

255 255 FT STRAND

270 270 FT HELIX

285 286 FT TURN

287 287 FT HELIX

306 297 FT HELIX

315 309 FT HELIX

319 316 FT STRAND

333 333 FT TURN

348 343 FT STRAND

354 353 FT TURN

362 356 FT TURN

364 363 FT HELIX

370 367 FT TURN

371 371 FT TURN

375 375 FT HELIX

387 387 FT TURN

399 399 FT TURN

Query Match 16.8%; Score 481; DB 1; Length 614;

Best Local Similarity 30.2%; Pred. No. 7.3e-27;

Matches 162; Conservative 78; Mismatches 177; Indels 120; Gaps 26;

Qy 35 AFLGIPFAEPVGNLRFKDPV---PVSGSLNQKFTSYGSCMQ-----QNPEGTFEEN 85

Db 62 AFLGIPFAEPVGNLRFKDPV---PVSGSLNQKFTSYGSCMQ-----QNPEGTFEEN 118

Qy 86 LGKTALDLVMSQKVFQAVLPQSEDCLTINVRPPGKTGAGANLPVMLWIFGGGFEGSPRTI 145

Db 119 PNREL-----SEDCLTINVRPPGKTGAGANLPVMLWIFGGGFEGSPRTI 161

Qy 146 FPPAQWVTKSVLMGKHI-----IHVAVNYRVASNGFLAGDDIKAEAGSGNAGLKQORLGM 199

Db 162 -----DVYDGRFLAQVEGAVLVMNRYGVTEGFLALPGSR-EAPGNVGLLDQRLAL 211



RESULT 14  
ACES LEPDE STANDARD; PRT; 629 AA.

AC Q27677;  
01-NOV-1997 (Rel. 35, Created)  
01-NOV-1997 (Rel. 35, Last sequence update)  
16-OCT-2001 (Rel. 40, Last annotation update)  
DE Acetylcholinesterase precursor (EC 3.1.1.7) (ACHE).  
OS Leptinotarsa decemlineata (Colorado potato beetle).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;  
OC Phytophaga; Chrysomeloidea; Chrysomelidae; Chrysomelinae;  
OC Chrysomelini; Leptinotarsa.  
OC NCBI\_TaxID=7539;  
RN [1]  
SEQUENCE FROM N.A.  
RP STRAIN=SS; TISSUE=Larva, and Pupae;  
RC MEDLINE=96137147; PubMed=8580913;  
RX Zhu K.Y., Clark J.M.;  
RT "Cloning and sequencing of a cDNA encoding acetylcholinesterase in  
RT Colorado potato beetle, Leptinotarsa decemlineata (Say).";  
RL Insect Biochem. Mol. Biol. 25:1129-1138(1995).  
CC -|- FUNCTION: Rapidly hydrolyzes choline released into the synapse.  
CC -|- CATALYTIC ACTIVITY: Acetylcholine + H(2)O = choline + acetate.  
CC -|- SUBCELLULAR LOCATION: Linked to the membrane of the neuronal  
CC cholinergic synapses by a GPI-anchor (By similarity).  
CC -|- PTM: The N-terminus is blocked.  
CC -|- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L41180; AAB00466.1; -;  
DR HSSP; P21836; 1MAA.  
DR InterPro; IPR002018; CarboxylesteraseB.  
DR InterPro; IPR000997; Cholinesterase.  
DR InterPro; IPR000379; Ser\_estrs.  
DR Pfam; PF00135; Coesterase; 1.  
DR PRINTS; PR00878; CHOLNESTRASE.  
DR PROSITE; PS00122; CARBOXYLESTERASE B\_1; 1.  
DR PROSITE; PS00941; CARBOXYLESTERASE B\_2; 1.  
DR Hydrolase; Serine esterase; Synapse; Membrane; Nerve; Signal;  
KW Neurotransmitter degradation; GPI-anchor; Glycoprotein.  
FT SIGNAL 1..29 POTENTIAL.  
FT CHAIN 30..629 ACETYLCHOLINESTERASE.  
FT ACT\_SITE 253 BY SIMILARITY.  
FT ACT\_SITE 382 BY SIMILARITY.  
FT ACT\_SITE 496 BY SIMILARITY.  
FT DISULFID 103 BY SIMILARITY.  
FT DISULFID 307 BY SIMILARITY.  
FT DISULFID 458 BY SIMILARITY.  
FT DISULFID 595 INTERCHAIN (BY SIMILARITY).  
FT CARBOHYD 125 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 308 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 418 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 509 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 629 AA; 71142 MW; 06556F833EB16C72 CRC64;  
Query Match 16.5%; Score 472; DB 1; Length 629;  
Best Local Similarity 28.3%; Pred. No. 3.3e-26;  
Matches 158; Conservative 79; Mismatches 199; Indels 122; Gaps 24;  
QY 16 TAKLANGDTITGLNAINEAFIGLFPAPPPVGNLRFKDPV---PYGSLNGQKFTSYGSPS 72  
DB 47 TSGLVRGYSKTVLGRVH-VFTGIPAKPPIEQLRKKFVPIDPWGILDATQPN---S 102  
QY 73 CMOQNPE---GTFENLKGKTLDLVMQSKVQFVLPDQSEDLITNVVRP-----P 119

RESULT 14  
ACES LEPDE STANDARD; PRT; 629 AA.

ID	ACES LEPDE	STANDARD;	PRT;	629 AA.
AC	Q27677;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Acetylcholinesterase precursor (EC 3.1.1.7) (ACHE).			
OS	Leptinotarsa decemlineata (Colorado potato beetle).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;			
OC	Phytophaga; Chrysomeloidea; Chrysomelidae; Chrysomelinae;			
OC	Chrysomelini; Leptinotarsa.			
OX	NCBI_TaxID=7539;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=SS; TISSUE=Larva, and Pupae;			
RC	MEDLINE=96137147; PubMed=8580913;			
RX	Zhu K.Y., Clark J.M.;			
RT	"Cloning and sequencing of a cDNA encoding acetylcholinesterase in			
RT	Colorado potato beetle, <i>Leptinotarsa decemlineata</i> (Say).";			
RL	Insect Biochem. Mol. Biol. 25:1129-1138 (1995).			
CC	-1- FUNCTION: Rapidly hydrolyzes choline released into the synapse.			
CC	-1- CATALYTIC ACTIVITY: Acetylcholine + H <sub>2</sub> O = choline + acetate.			
CC	-1- SUBCELLULAR LOCATION: Linked to the membrane of the neuronal			
CC	cholinergic synapses by a GPI-anchor (By similarity).			
CC	-1- PTM: The N-terminus is blocked.			
CC	-1- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; L41180; AAB00466.1; -			
DR	HSSP; P21836; 1MAA.			
DR	InterPro; IPR002018; CarboxylesteraseB.			
DR	InterPro; IPR000997; Cholinesterase.			
DR	InterPro; IPR000379; Ser_estrs.			
DR	Pfam; PF00135; Coesterase; 1.			
DR	PRINTS; PR00878; CHOLNESTRASE.			
DR	PROSITE; PS00122; CARBOXYLESTERASE B 1; 1.			
DR	PROSITE; PS00941; CARBOXYLESTERASE B 2; 1.			
KW	Hydrolase; Serine esterase; Synapse; Membrane; Nerve; Signal;			
KW	Neurotransmitter degradation; GPI-anchor; Glycoprotein.			
FT	SIGNAL 1 29			
FT	POTENTIAL.			
FT	CHAIN 30 629			
FT	ACETYLCHOLINESTERASE.			
FT	ACT_SITE 253 253			
FT	BY SIMILARITY.			
FT	ACT_SITE 382 382			
FT	BY SIMILARITY.			
FT	ACT_SITE 496 496			
FT	BY SIMILARITY.			
FT	DISULFID 103 130			
FT	BY SIMILARITY.			
FT	DISULFID 307 322			
FT	BY SIMILARITY.			
FT	DISULFID 458 574			
FT	INTERCHAIN (BY SIMILARITY).			
FT	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD 125 125			
FT	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD 308 308			
FT	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD 418 418			
FT	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD 509 509			
FT	N-LINKED (GLCNAC. .) (POTENTIAL).			
SQ	SEQUENCE 629 AA; 71142 MW; 06556F833EB16C72 CRC64;			

Query Match 16.5%; Score 472; DB 1; Length 629;  
Best Local Similarity 28.3%; Pred. No. 3.3e-26;  
Matches 158; Conservative 79; Mismatches 199; Indels 122; Gaps 24;

ID	ACES LEPDE	STANDARD;	PRT;	629 AA.
QY	16 TAKLANGDTITGLNAINEAFIGFIPFAPPPVGNLRFKDPV---PYGSLNGQKFTSYGSPS 72			
DB	47 TSGLVRGYSKTVLGRVH-VFTGIPAKPPIEQLRKKFVPIDPWGILDATQPN---S 102			
QY	73 CMOQNPE---GTFENLKGALDLVMQSKVFQAVLPDQSEDLITNVVRP-----P 119			

```

ranging from monomers and dimers to collagen-tailed and
hydrophobic-tailed forms, in which catalytic tetramers are
associated with anchoring proteins that attach them to the basal
lamina or to cell membranes. In the collagen-tailed forms, isoform
T subunits are associated with a specific collagen. COLQ, which
triggers the formation of isoform T tetramers, from monomers and
dimers.
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=T;
IsoId=P23795-1; Sequence=Displayed;
Name=H;
IsoId=P23795-2; Sequence=VSP_001455;
-!- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).

```

```

72 QY 72 SCHQ-----ONPEGTEENLGKTDALDMQSKVQAVLPQSEDECLTINVVRPPGTAGA 123
Db 99 -CYQYVDTLTPGFEGETEMPNREL:|||||SEDECLYN-VMTYPPRESS 140

126 QY 126 NLPVLMWIFGCGFEIGSPITFPFPAQMTKSVLMGKHI-----IHVAVNYRVASWGFLAG 179
Db 141 PTPVLVWVYGGFFYSGASSL:|||||DYGDRFLQTQABGTVLVSNMYRVGAFGLAL 191

180 QY 180 DDIAKEGSGNAGLKQDLGMQWADNIAFGGDPSPKVTIFGESAGSMVLCILLIWNDDGN 239
Db 192 PGSR-EAPGNVGLLDQRLALQWQENVAAFGGDPTSVTLFGESAGAASVGMHLL-----244

240 QY 240 TYGKGPFRAGIMOSGAMVSDP--VDGTYGNEIYDLFVSSAGC-----GSASDKLAC 290
Db 245 SPFSRGLFHRVAVLQSGA--PNGPWATVGVGEARRRATLLARLVGCPPCGAGNDTEL VAC 302

291 QY 291 LRSASSDTLLDATNTPGFLAYSLSRLSYLPRPDGKNITDMYKLVDRGKYASVPVILGD 350
Db 303 LRAPAOQLVD--HEWRVLQESVFRPSFVPVVDGDFLSLDPREALINAGDFHGLQVLGV 360

351 QY 351 QNDRGTTIF-----GLSSLN--VTTNQAARYFKQSFIHAGDAEIDTLMAY-----PQD 397
Db 361 VKDEGSYFLVYGAPGFKDNESLISRAQFLAGVRVGPQASDLAAEAIVLHYTLWHPED 420

398 QY 398 ---ITQSGPDPDTGVNLALTPOFKKISAVLGLDAFIHARRYFLNHFQGGTKYSFL-----SK 450
Db 421 PARLREALSDVVGDNHVVCP-----VAQLAGSLA-----AQGARVYAITFEHRAS 465

451 QY 451 QLSGLPIWGTGFHANDIWDQYLLG--SGSVYNNAFIAFATDL-----DPNTA 496
Db 466 TLSWFLWVGVPHGVEI---EFIFGLPLEPSLNYITEERTFAQRLMYWANPARTGDPNDP 522

497 QY 497 --GLLVNWPKYTSSSQ 510
Db 523 RDPKAPQWPPYTAGAQ 538

```

Search completed: July 29, 2004, 10:49:33  
Job time : 14 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 10:47:09 ; Search time 21 Seconds  
(without alignments)  
2505.560 Million cell updates/sec

Title: US-09-943-857-4  
Perfect score: 2864  
Sequence: 1 SMNSRGPAGRLGVPKALIA.....DNFRTAGYDALMTNPSPFFV 547

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2779	97.0	549	JN0551	triacylglycerol li
2	2775	96.9	534	S41735	cholesterol estera
3	2513	87.7	549	S23448	triacylglycerol li
4	2431	84.9	549	JN0553	triacylglycerol li
5	2415	84.3	538	S05684	lipase I precursor
6	2366	82.6	549	JN0552	triacylglycerol li
7	2314	80.8	548	S32615	triacylglycerol li
8	1095	38.2	544	S41091	triacylglycerol li
9	1092	38.1	563	1 ACGUGC	triacylglycerol li
10	1089	38.0	544	S41093	triacylglycerol li
11	1086	37.9	544	S59958	triacylglycerol li
12	1084	37.8	544	S41092	triacylglycerol li
13	1081	37.7	544	S41096	triacylglycerol li
14	1081	37.7	563	S59957	triacylglycerol li
15	1078	37.6	563	1 PN0493	triacylglycerol li
16	1075	37.5	544	S41090	triacylglycerol li
17	1070	37.4	544	S41095	triacylglycerol li
18	1069	37.3	544	S41094	triacylglycerol li
19	491.5	17.2	596	1 ACRYE	acetylcholinestera
20	489	17.1	599	1 A38868	acetylcholinestera
21	481	16.8	614	2 JH0314	cholinesterase (EC
22	476	16.6	603	2 S70849	cholinesterase (EC
23	469	16.4	614	2 JH0811	acetylcholinestera
24	461	16.1	583	2 S10712	acetylcholinestera
25	459	16.0	581	2 C39768	cholinesterase (EC
26	452	15.8	614	2 A39256	acetylcholinestera
27	451	15.7	584	2 T27009	hypothetical prote
28	451	15.7	602	1 ACHU	cholinesterase (EC
29	447.5	15.6	584	2 S48724	acetylcholinestera

30	442	15.4	607	2	T42399	acetylcholinestera
31	430.5	15.0	489	2	B69680	para-nitrobenzyl e
32	427.5	14.9	620	2	A54413	acetylcholinestera
33	423	14.8	532	2	A34329	60K esterase (EC 3
34	418	14.6	540	2	A75250	carboxylesterase,
35	411.5	14.4	597	2	A33668	sterol esterase (E
36	410	14.3	565	2	S10367	carboxylesterase (
37	409	14.3	549	2	JX0054	carboxylesterase (
38	408.5	14.3	561	2	S47655	carboxylesterase (
39	407	14.2	637	2	S66236	acetylcholinestera
40	405.5	14.2	547	2	T29717	hypothetical prote
41	404.5	14.1	664	2	JC7990	acetylcholinestera
42	404	14.1	691	2	JE0150	acetylcholinestera
43	402	14.0	561	2	S62788	carboxylesterase (
44	401	14.0	554	2	A39060	carboxylesterase (
45	401	14.0	746	2	A25363	acetylcholinestera

ALIGNMENTS

RESULT 1

JN0551  
triacylglycerol lipase (EC 3.1.1.3) 3 precursor - yeast (Candida rugosa)  
N;Alternate names: lipase; triacylglycerol acylhydrolase  
C;Species: Candida rugosa  
C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 07-May-1999  
C;Accession: JN0551; S29633  
R;Lotii, M.; Grandori, R.; Fusetti, F.; Longhi, S.; Brocca, S.; Tramontano, A.; Alberghi  
Gene 124, 45-55, 1993  
A;Title: Cloning and analysis of Candida cylindracea lipase sequences.  
A;Reference number: JN0551; MUID:93178975; PMID:8440480  
A;Accession: JN0551  
A;Molecule type: DNA  
A;Residues: 1-549 <LOT>  
A;Cross-references: EMBL:X66008  
A;Note: The source is designated as Candida cylindracea  
R;Rua, M.L.; Diaz-Maurino, T.; Fernandez, V.M.; Otero, C.; Ballesteros, A.  
Biochim. Biophys. Acta 181, 181-189, 1993  
A;Title: Purification and characterization of two distinct lipases from Candida cylindra  
A;Reference number: S29633; MUID:93152596; PMID:8427877  
A;Accession: S29633  
A;Molecule type: protein  
A;Residues: 16-25 <RUA>  
A;Note: the source is designated as Candida cylindracea  
C;Genetics:  
A;Gene: LIP3  
C;Function:  
Description: hydrolyzes triacylglycerols into fatty acids and glycerol  
C;Superfamily: triacylglycerol lipase  
C;Keywords: carboxylic ester hydrolase; disulfide bond; glycoprotein  
F;1-15/Domain: signal sequence #status predicted <SIG>  
F;16-549/Product: triacylglycerol lipase 3 #status experimental <MAT>  
F;222-226/Region: interfacial lipid recognition (GXSG) motif  
F;75-112/Disulfide bonds: #status predicted  
F;224/Active site: Ser #status predicted  
F;329,366/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 97.0%; Score 2779; DB 2; Length 549;  
Best Local Similarity 98.5%; Pred. No. 1.5e-192;  
Matches 529; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY	11	LGSVPTAKLANGDTITGLNAINEAFGLPPFAEPVGNLRFKDPVPYSGSLNGKQKTSYG	70
DB	13	VAAAPTAKLANGDTITGLNAINEAFGLPPFAEPVGNLRFKDPVPYSGSLNGKQKTSYG	72
QY	71	PSCMQONPGTEENLGKTLDLVMQSFQAVLPQSEDLTINVVRPPTKAGANLPVM	130
DB	73	PSCMQONPGTEENLGKTLDLVMQSFQAVLPQSEDLTINVVRPPTKAGANLPVM	132
QY	131	LWIFGGFPIGSPITPPPAQMVKSVLMGKHHIVAVNRYRVSWSGFLAGDDIKAECSGNA	190
DB	133	LWIFGGFPIGSPITPPPAQMVKSVLMGKHHIVAVNRYRVSWSGFLAGDDIKAECSGNA	192

Qy	191	GLKDQRLGMQWADNIAGFGGDPKVTIIFGSAGSMSVLCHLIWMDGNTYKGKPLFRAG	250
Db	193	GLKDQRLGMQWADNIAGFGGDPKVTIIFGSAGSMSVLCHLIWMDGNTYKGKPLFRAG	252
Qy	251	IMQSGAMVPSDPVDGTYGNEIYDLFVSSAGCGSASDKLACLSASDITLLDATNNTPGFL	310
Db	253	IMQSGAMVPSDPVDGTYGNEIYDLFVSSAGCGSASDKLACLSASDITLLDATNNTPGFL	312
Qy	311	AYSSLRLSYLPDPDGKNIITDDMYKLVRDGKYASVPVIIGDQNDGFTIFGLSSINVTNNAQ	370
Db	313	AYSSLRLSYLPDPDGKNIITDDMYKLVRDGKYASVPVIIGDQNDGFTIFGLSSINVTNNAQ	372
Qy	371	ARAYFKQSFIHASDAEIDITLMAAYPQDITQGSPPDTGVNALNTPQFKRISAVLGDLAFTH	430
Db	373	ARAYFKQSFIHASDAEIDITLMAAYPQDITQGSPPDTGIFNAITPQFKRISAVLGDLAFTH	432
Qy	431	ARRYFLNHFQCGTKYKSFUSKOLSGLPIMGTFTHANDIWMQDYLLGGSGSVYNNNAFTAFATD	490
Db	433	ARRYFLNHFQCGTKYKSFUSKOLSGLPIMGTFTHANDIWMQDYLLGGSGSVYNNNAFTAFATD	492
Qy	491	LDPNTAGLLVNWPKYTSSTSSQSGNNLMINALGLYTGKDNFRTAGYDALMTNPSSPFV	547
Db	493	LDPNTAGLLVNWPKYTSSTSSQSGNNLMINALGLYTGKDNFRTAGYDALMTNPSSPFV	549
RESULT 2			
S41735			
cholesterol esterase - yeast (Candida rugosa)			

```

Db      362  FKQSFIIHSDAIEDITLMAYPQDITQGSFFDTGTFNALTPQPKRTISAVILGDLIAFHARRY 421
Qy      435  FLNFHQGQTKYSFSLSKQLSGLPIMGTFTHANDIVWQDYLLGSGSVIYNNAFAFAFATLDLDN 494
Db      422  FLNFHQGQTKYSFSLSKQLSGLPIMGTFTHANDIVWQDYLLGSGSVIYNNAFAFAFATLDLDN 481
Qy      495  TAGLLVNWPKYTSSTSSQSGNNLMMLNGLYTGKGNFRTAGYDALMTPNPSFFV 547
Db      482  TAGLLVNWPKYTSSTSSQSGNNLMMLNGLYTGKGNFRTAGYDALMTPNPSFFV 534

RESULT 3
S23448
C:Species: Candida rugosa
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 03-May-1996
C:Accession: S23448
R;Longhi, S.; Fusetti, F.; Grandori, R.; Lotti, M.; Vanoni, M.; Alberghina, L.
Biochim. Biophys. Acta 1131, 227-232, 1992
A:Title: Cloning and nucleotide sequences of two lipase genes from Candida cylindracea
A:Reference number: S23448; MUID:92305068; PMID:1610906
A:Accession: S23448
A:Molecule type: DNA
A:Residues: 1-549 <LON>
A:Cross-references: EMBL:X64703
A>Note: the source is designated as Candida cylindracea
C:Superfamily: triacylglycerol lipase
C:Keywords: carboxylic ester hydrolase

```



N:Alternate names: lipase; triacylglycerol acylhydrolase  
 C:Species: Candida rugosa  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 07-May-1999  
 C:Accession: JN0553; S29634  
 R:Lotti, M.; Grandori, R.; Fusetti, F.; Longhi, S.; Brocca, S.; Tramontano, A.; Alberghi  
 Gene 124, 45-55, 1993  
 A:Title: Cloning and analysis of Candida cylindracea lipase sequences.  
 A:Reference number: JN0551; MUID:93178975; PMID:8440480  
 A:Accession: JN0553  
 A:Molecule type: DNA  
 A:Residues: 1-549 <LOT>  
 A:CROSS-references: EMBL:X66008  
 A:Note: the source is designated as Candida cylindracea  
 R:Rua, M.L.; Diaz-Maurino, T.; Fernandez, V.M.; Otero, C.; Ballesteros, A.  
 Biochim. Biophys. Acta 1156, 181-189, 1993  
 A:Title: Purification and characterization of two distinct lipases from Candida cylindracea  
 A:Reference number: S29633; MUID:93152596; PMID:8427877  
 A:Accession: S29634  
 A:Molecule type: protein  
 A:Residues: 16-25 <RUA>  
 A:Note: the source is designated as Candida cylindracea  
 C:Genetics:  
 A:Gene: LIPS  
 C:Function:  
 A:Description: hydrolyzes triacylglycerols into fatty acids and glycerol  
 C:Superfamily: triacylglycerol lipase  
 C:Keywords: carboxylic ester hydrolase; disulfide bond; glycoprotein  
 F:1-15/Domain: signal sequence #status predicted <SIG>  
 F:16-549/Product: triacylglycerol lipase 5 #status experimental <MAT>  
 F:222-226/Region: interfacial lipid recognition (GXSGX) motif  
 F:75-112/Disulfide bond: #status predicted  
 F:224/Active site: Ser #status predicted  
 F:329,366/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 84.9%; Score 2431; DB 2; Length 549;  
 Best Local Similarity 85.5%; Pred. No. 1.9e-167;  
 Matches 459; Conservative 32; Mismatches 46; Indels 0; Gaps 0;

QY	11	LGSVPTAKLANGDTITGLNAINEAFGLIPPAEPVGNLRFKDPVYSGSLNGKQTSYG	70
DB	13	VAAAPTATLANGDTITGLNAINEAFGLIPPAEPVGNLRFKDPVYSGSLNGKQTSYAG	72
QY	71	PSCMQONPEGTPEENLGKTDALVMSQKQVAVLPQSEDCLTINVRPPTKAGANLPVM	130
DB	73	PSCMQONPEGTPEENLPKVALDLVMSQKQVAVLPQSEDCLTINVRPPTKAGANLPVM	132
QY	131	LWIFGGFEGSPTIPPPAQMTKSVLMGKHIIHVAVNRYKVASWGLAGDDIIKAEGSGNA	190
DB	133	LWIFGGFEGSPTIPPPAQMTKSVLMGKPIIHVAVNRYKVASWGLAGDDIIKAEGSSNA	192
QY	191	GLKQDRGLQMWADNITAGFGDPSKVTIFGESAGSMVLCHLIWNDDNTYKGPPLFRAG	250
DB	193	GLKQDRGLQMWADNITAGFGDPSKVTIFGESAGSMVLCHLIWNDDNTYKGPPLFRAG	252
QY	251	IMQSGAMVPSPDVGDTYGNIEYDLFVSSAGCGSASDKLACLRASSDITLLDATNTPGFL	310
DB	253	IMQSGAMVPSPDVGDTYGTQIYDLFVASTGSSASDKLACLRASSDITLLDATNTPGFL	312
QY	311	AYSSLRLSYLPRDGNITDDMYKLVDRDGKYASVPVVIIGQNDDEGTIFGLSSLNVTNAQ	370
DB	313	SYTSLRLSYLPRDGNITDDMYKLVDRDGKYASVPVVIIGQNDDEGTIFGLSSLNVTTEAD	372
QY	371	ARAYFKQSFTHASDAEIDTILMAAYPDITQSGPDTGVNALTPQPKRISAVLGDLAFIH	430
DB	373	AEALYRKSFTHTADADITALKAAVPSDVTQSGPDTGILNALTPQPKRINAVLGDLFTTL	432
QY	431	ARRYFLNHFGGTYKYSFLSKQLSGLPIMGTFFHNDIVWQDYLLGSGSVIYNNAFIATD	490
DB	433	SRRYFLNHYTGPKYSFLSKQLSGLPILGTFFHNDIVWQDYLLGSGSVIYNNAFIATD	492
QY	491	LDPNNTAGLLVWPKYTSQSGNNLMINALGLYTGKDNFTAGYDALMTNPSPFFV	547
DB	493	LDPNNTAGLLVQPKTSQSGAGDNLQISALGLYTGKDNFTAGYDALMTNPSPFFV	549

## RESULT 5

S05684  
 lipase I precursor - yeast (Candida rugosa) (fragment)  
 C:Species: Candida rugosa  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 13-Sep-1998  
 C:Accession: S05684  
 R:Kawaguchi, Y.; Honda, H.; Taniguchi-Morimura, J.; Iwasaki, S.  
 Nature 341, 164-166, 1989  
 A:Title: The codon CUG is read as serine in an asporogenic yeast Candida cylindracea.  
 A:Reference number: S05684; MUID:89384874; PMID:2506450  
 A:Accession: S05684  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-538 <KAW>  
 A:Note: part of this sequence, including the amino end of the mature protein, was confirmed as Candida cylindracea  
 C:Superfamily: triacylglycerol lipase  
 F:1-4/Domain: signal sequence (fragment) #status predicted <SIG>  
 F:5-538/Product: lipase I #status experimental <MAT>

Query Match 84.3%; Score 2415; DB 2; Length 538;  
 Best Local Similarity 84.5%; Pred. No. 2.7e-166;  
 Matches 454; Conservative 32; Mismatches 51; Indels 0; Gaps 0;

QY	11	LGSVPTAKLANGDTITGLNAINEAFGLIPPAEPVGNLRFKDPVYSGSLNGKQTSYG	70
DB	2	VAAAPTATLANGDTITGLNAINEAFGLIPPAEPVGNLRFKDPVYSGSLNGKQTSYG	61
QY	71	PSCMQONPEGTPEENLGKTDALVMSQKQVAVLPQSEDCLTINVRPPTKAGANLPVM	130
DB	62	PLCMQONPEGTPEENLPKAAOLDVMSQKQVAVLPQSEDCLTINVRPPTKAGANLPVM	121
QY	131	LWIFGGFEGSPTIPPPAQMTKSVLMGKHIIHVAVNRYKVASWGLAGDDIIKAEGSGNA	190
DB	122	LWIFGGFEGSPTIPPPAQMTKSVLMGKPIIHVAVNRYKVASWGLAGDDIIKAEGSANA	181
QY	191	GLKQDRGLQMWADNITAGFGDPSKVTIFGESAGSMVLCHLIWNDDNTYKGPPLFRAG	250
DB	182	GLKQDRGLQMWADNITAGFGDPSKVTIFGESAGSMVLCHLIWNDDNTYKGPPLFRAG	241
QY	251	IMQSGAMVPSPDVGDTYGNIEYDLFVSSAGCGSASDKLACLRASSDITLLDATNTPGFL	310
DB	242	IMQSGAMVPSPDVGDTYGNIEYDLFVSSAGCGSASDKLACLRASSDITLLDATNTPGFL	301
QY	311	AYSSLRLSYLPRDGNITDDMYKLVDRDGKYASVPVVIIGQNDDEGTIFGLSSLNVTNAQ	370
DB	302	AYSSLRLSYLPRDGNITDDMYKLVDRDGKYASVPVVIIGQNDDEGTIFGLSSLNVTDAQ	361
QY	371	ARAYFKQSFTHASDAEIDTILMAAYPDITQSGPDTGVNALTPQPKRISAVLGDLAFIH	430
DB	362	ARAYFKQSFTHASDAEIDTILMAAYPDITQSGPDTGVNALTPQPKRISAVLGDLGFTL	421
QY	431	ARRYFLNHFGGTYKYSFLSKQLSGLPIMGTFFHNDIVWQDYLLGSGSVIYNNAFIATD	490
DB	422	ARRYFLNHYTGPKYSFLSKQLSGLPILGTFFHNDIVWQDYLLGSGSVIYNNAFIATD	481
QY	491	LDPNNTAGLLVWPKYTSQSGNNLMINALGLYTGKDNFTAGYDALMTNPSPFFV	547
DB	482	LDPNNTAGLLVQPKTSQSGAGDNLQISALGLYTGKDNFTAGYDALMTNPSPFFV	538

## RESULT 6

JN0552  
 triacylglycerol lipase (EC 3.1.1.3) 4 precursor - yeast (Candida rugosa)  
 N:Alternate names: lipase; triacylglycerol acylhydrolase  
 C:Species: Candida rugosa  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 07-May-1999  
 C:Accession: JN0552  
 R:Lotti, M.; Grandori, R.; Fusetti, F.; Longhi, S.; Brocca, S.; Tramontano, A.; Alberghi  
 Gene 124, 45-55, 1993  
 A:Title: Cloning and analysis of Candida cylindracea lipase sequences.



Best Local Similarity 43.5%; Pred. No. 4.1e-71;  
Matches 238; Conservative 72; Mismatches 193; Indels 44; Gaps 9;

QY 15 PTAFLANGTITGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGSLNGQKTSYGPSCM 74  
Db 3 PTAFLANGNEISGVLEKGVDTFKGIPFADPPVGDLPFKHPQPTGSGYQGLKANDFSSACM 62  
QY 75 QNPEGTTFE-----ENLCKTALDLVMSQKVFQAVLPQSEDCLTINVRPPGT 121  
Db 63 QLDPGNAISLLDKVGLGKILPDNLRGPLYDMA-----QGSVSMNEDCLYLVFPAGT 116  
QY 122 KAGANLPVMLWIFGGGFEIGSPITFPPAQMTKSVLMGKHIIHVAVNYRVASVGLAGDD 181  
Db 117 KPDALPVMVWIYGGAFVFGSSASYPGVVKSVEVMGQPVVFSINVRTGPGYFLGGDA 176  
QY 182 IKAEGSGNAGLKQRLGMOWADNIAGFGDPSKVTIFGESAGSMVYLCHLIWDDGNTY 241  
Db 177 ITAEGSTNAGLHQRLGLEWVSDNIANFGGDPKVMIFGESAGMSVAHQLVAYGDDNTY 236  
QY 242 KGKPLFRAGIMQSGAMVPSDPVDTGYNEIYDLFVSSAGC---GSASDKLACLRSSSDT 298  
Db 237 NGQLFHSAILOQSGGFLPYFDSTVSGPSAYSRFAQYAGCDTSASDNDTLACLRSSDV 296  
QY 299 LLDATNN-----TPGFLAYSLSRLSYLPRPDGKNITDDMYKLVRDGGKYASVPIIG 349  
Db 297 LHSQNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELVRSGRYAKVPIIG 349  
QY 350 QNDDEGTIFGLSLNVTNAQARAYFKQSFHSDAIEDITLMAAYPQDITQSGPFTGVL 409  
Db 350 NOEDEGITLAPVAINATTTTPHVKKWLKICQSDASDLSRVLSPGSESGSPFTGIL 409  
QY 410 NALTPOFKRISAVGLDIAFHARRYFLNHFGQGTYSFLSKQLSGL-PIMGTFHANDIVM 468  
Db 410 NALTPOFKRIAAITFDLLFQSPRRVMLNATKDVNRVTYLTATLQNLNVPFLTFHSGDLLF 469  
QY 469 QDYL-LGSGSVIYNNAFIATPDLDPNTAGLLVNWPKYTSSSQSGNNLMINALGTYGK 527  
Db 470 QYVVDLGPSA-YRRYFISFANHHDPNVTNLQWDMYT---DAGKEMLIQHMGNSMRT 525  
QY 528 DNFRTAG 534  
Db 526 DDFRIEG 532

RESULT 9  
ACGUCG  
triacylglycerol lipase (EC 3.1.1.3) I precursor - yeast (Geotrichum candidum)  
N;Alternate names: lipase  
C;Species: Geotrichum candidum  
C;Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 31-Mar-2000  
C;Accession: P0492; JQ0022  
R;Nagao, T.; Shimada, Y.; Sugihara, A.; Tominaga, Y.  
J. Biochem. 113, 776-780, 1993  
A;Title: Cloning and sequencing of two chromosomal lipase genes from Geotrichum candidum  
A;Reference number: P0492; MUID:93380907; PMID:8370674  
A;Accession: P0492  
A;Molecule type: DNA  
A;Residues: 1-563 <NAG>  
A;Note: The translation of residues 31-550 and the corresponding nucleotide sequence are  
R;Shimada, Y.; Sugihara, A.; Tominaga, Y.; Iizumi, T.; Tsunawawa, S.  
J. Biochem. 106, 383-388, 1989  
A;Title: cDNA molecular cloning of Geotrichum candidum lipase.  
A;Reference number: JQ0022; MUID:90110016; PMID:2481674  
A;Accession: JQ0022  
A;Molecule type: mRNA  
A;Residues: 1-563 <SHI>  
A;Experimental source: strain ATCC 34614  
A;Note: sequences of several small peptides were also determined  
C;Comment: The extracellular lipase produced by Geotrichum candidum hydrolyzes all ester  
C;Genetics:  
A;Gene: lipi  
C;Superfamily: cholinesterase; cholinesterase homology  
C;Keywords: carboxylic ester hydrolase; glycoprotein; pyroglutamic acid

F1-19/Domain: signal sequence #status predicted <SIG>  
F120-563/Product: triacylglycerol lipase #status experimental <MAT>  
F143-560/Domain: cholinesterase homology <CHE>  
F1234-238/Region: interfacial lipid recognition (GXSG) motif  
F120/Modified site: pyrolidone carboxylic acid (Gln) (in mature form) #status experimen  
F180-124,295-307/Disulfide bonds: #status predicted  
F1236/Active site: Ser #status predicted  
F1302,383/Binding site: carbohydtrate (Asn) (covalent) #status predicted

Query Match 38.1%; Score 1092; DB 1; Length 563;  
Best Local Similarity 42.9%; Pred. No. 7.1e-71;  
Matches 237; Conservative 72; Mismatches 201; Indels 42; Gaps 8;

QY 9 GRIGSVPTAKLANGDTITGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGSLNGQKFTS 68  
Db 16 GTLAQAPTAVLANGNEISGVLEKGVDTFKGIPFADPPVGDLPFKHPQPTGSGYQGLKAND 75  
QY 69 YGSPCMQONPEGTTFE-----ENLCKTALDLVMSQKVFQAVLPQSEDCLTINV 115  
Db 76 FSSACMLQDFGNAISLLDKVGLGKILPDNLRGPLYDMA-----QGSVSMNEDCLYLVN 129  
QY 116 VPPGPTKAGANLPVMLWIFGGGFEIGSPITFPPAQMTKSVLMGKHIIHVAVNYRVASWG 175  
Db 130 FRPAGTKPDAKLPVMVWIYGGAFVFGSSASYPGVVKSVEVMGQPVVFSINVRTGPGY 189  
QY 176 FLAGDDIKAEGSGNAGLKQRLGMOWADNIAGFGDPSKVTIFGESAGSMVYLCHLIWN 235  
Db 190 FLGGDAITAEAGNTNAGLHQRLGLEWVSDNIANFGGDPKVMIFGESAGMSVAHQLVAY 249  
QY 236 DGNTYKGPPLFRAGIMQSGAMVPSDPVDTGYNEIYDLFVSSAGC-GSASDK--LACLR 292  
Db 250 GGDNTYNGQLFHSAILOQSGGFLPYFDSTVSGPSAYSRFAQYAGCDASAGDNETLACLR 309  
QY 293 SASSDTLLDATNN-----TPGFLAYSLSRLSYLPRPDGKNITDDMYKLVRDGGKYAS 343  
Db 310 SKSDSVLHSAQNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELVRSGEYAK 362  
QY 344 VPVIIGQNDDEGTIFGLSLNVTNAQARAYFKQSFHSDAIEDITLMAAYPQDITQSGP 403  
Db 363 VPVIIGQNDDEGTILAPVAINATTTTPHVKKWLKICSEASDASLDRVLSPGSESGAP 422  
QY 404 FDTGVLNALTPOFKRISAVGLDIAFHARRYFLNHFGQGTYSFLSKQLSGL-PIMGTFH 462  
Db 423 FRTGILNALTPOFKRIAAITFDLLFQSPRRVMLNATKDVNRVTYLTATLQNLNVPFLTFH 482  
QY 463 ANDIVMQDYLGGSGSVIYNNAFIATPDLDPNTAGLLVNWPKYTSSSQSGNNLMINALG 522  
Db 483 GSDLLFYVYAGPWSSSAYRRYFISFANHHDPNVTNLQWDMYT---DSGKEMLIQHMG 539  
QY 523 LYTGKDNFRTAG 534  
Db 540 NSMRTDDFRIEG 551

RESULT 10  
S41093  
triacylglycerol lipase (EC 3.1.1.3) I precursor - yeast (Geotrichum candidum) (strain NF  
C;Species: Geotrichum candidum  
A;Variety: NRRL Y-553  
C;Date: 19-Mar-1997 #sequence\_revision 05-Feb-1999 #text\_change 18-Jun-1999  
C;Accession: S41093  
R;Bertolini, M.C.; Laramee, L.; Thomas, D.Y.; Cygler, M.; Schrag, J.D.; Vernet, T.  
Eur. J. Biochem. 219, 119-125, 1994  
A;Title: Polymorphism in the lipase genes of Geotrichum candidum strains.  
A;Reference number: S41090; MUID:94139683; PMID:8306978  
A;Accession: S41093  
A;Status: nucleic acid sequence not shown; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-544 <BER>  
A;Cross-references: GB:U02387; NID:9406507; PIDN:AAA03425.1; PID:9406508  
A;Experimental source: NRRL Y-553  
A;Note: only the translation of the mature protein is shown  
C;Genetics:

A:Gene: lipi  
A:Function:  
C:Description: hydrolyzes triacylglycerols into fatty acids and glycerol  
C:Superfamily: cholinesterase; cholinesterase homology; lipid hydrolysis; pyroglutamic acid  
C:Keywords: carboxylic ester hydrolase; glycoprotein; lipid hydrolysis; pyroglutamic acid  
F:24-541/Domain: cholinesterase homology <CHE>  
F:215-219/Region: interfacial lipid recognition (GXSGX) motif  
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status predicted  
F:61-105,276-288/disulfide bonds: #status predicted  
F:217/Active site: Ser #status predicted  
F:283,364/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 38.0%; Score 1089; DB 2; Length 544;  
Best Local Similarity 43.5%; Pred. No. 1.1e-70;  
Matches 238; Conservative 73; Mismatches 192; Indels 44; Gaps 10;

QY 15 PTAKLANGDTITGLNAIINEAFGLGIPFAEPVGNLRFKDPVPYSGSLNGQKFTSYGPGSCM 74  
DB 3 PTAVLNGNEISGLGVKVDTFKGIIPADPPVGLDRFKHPQFTGSGYQGLKANDFSACM 62  
QY 75 QONPEGTPE-----ENLGKLTALDLVMQSKVFAVLPOSEDCLTINVRPPGT 121  
DB 63 QLDPGNAISLWLDKVVGLGKILPDNLRGPLYDMA-----QGSVSNNEDECLINVRPAGT 116  
QY 122 KAGANLPVWLMTFGGFGPIGSGPTIPPPAQMVTKSLVLMGKHIHVAVNYRVASWGLAGDD 181  
DB 117 KPDAKLPVWMIYGGAFVFGSSASYPGNGYVKESVEMGQPVVFSINRTGPGYFLGGDA 176  
QY 192 IKAGSGNAGLKQDLQGLGMQVADNIAGFGDPSKVTIFGESAGSMVSLCHLIWNGDNTY 241  
DB 177 ITAEGNTNAGLHDKQGLGWSDNIANFGGDPDKVMIIFGESAGSMVAHQIYVAYGDDNTY 236  
QY 242 KGKPLFRAGIMQSGAMVPSDPVDGTGYGNEIYDLFVSSAGC--GSASDK--LACLRASDDT 298  
DB 237 NGKKLFHSAIILQSGGLPYFDSTSVGPESASRFAQYAGCDASDNETLACLRKSSDV 296  
QY 299 LLDATNN-----TPGFLAYSSLRLSYLPRPGKNITDDMYKLVRDGVKVASVPVITG 349  
DB 297 LHSAQNSYDLKDLGLLQFLGFG-----PRPDGNIIPDAAYELYSRGYAKVPYITG 349  
QY 350 DONDEGTIFGLSNLNTTNAQARAYFKOSFIHASDAEITDMAAYPQDITOGSPDITGV 409  
DB 350 NQDEGTILAPVAINATTPHVKWLVKICSEASDASLDRLSLYPSGSEGAERTGIL 409  
QY 410 NALTPOPKRISAVLGLDLAFIHARRYFLNHFGOGTKYSPLSKQLSGL--PIMGTFFHNDIVW 468  
DB 410 NALTPOPKRIAAITDILLFQSPRRVNLNATKDVNRTYLATQLHNLVFLGTFFHGSDDL 469  
QY 469 QDYL-LGSGSVIYNNAFIATPDTLDPNTAGLLVNWPKYTSQSGNNLMNALGLYTGK 527  
DB 470 QYYVDLGPSSA--YRRYFISFANHHDPNVGTLNQWDMYT---DAGKEMLIQHMVGNMRT 525  
QY 528 DNERTAG 534  
DB 526 DDFRIEG 532

RESULT 11  
S59958  
triacylglycerol lipase (EC 3.1.1.3) I - yeast (Geotrichum candidum) (strain CBS 178.71)  
C:Species: Geotrichum candidum  
A:Variety: strain CBS 178.71  
C:Date: 09-Mar-1996 #sequence\_revision 19-Apr-1996 #text\_change 21-Jan-2000  
C:Accession: S59958; S49236  
R:Phillips, A.; Pretorius, G.H.J.; van Rensburg, H.G.  
Biochem. Biophys. Acta 1252, 305-311, 1995  
A:Title: Molecular characterization of a Galactomyces geotrichum lipase, another member  
A:Reference number: S59957; MUID:96049515; PMID:7578238  
A:Accession: S59958  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-544 <PRT>  
A:Cross-references: EMBL:X81656

A:Experimental source: strain CBS 178.71  
R:Phillips, A.; Pretorius, G.H.J.; van Rensburg, H.G.  
submitted to the EMBL Data Library, September 1994  
A:Reference number: S49236  
A:Accession: S49236  
A:Molecule type: DNA  
A:Residues: 1-82, 'L', '84-239, 'Q', '241-383, 'Y', '385-544 <PHW>  
A:Cross-references: EMBL:X81656; NID:g547503; PIDN:CAA57316.1; PID:g547504  
C:Genetics:  
C:Gene: lipi  
C:Superfamily: cholinesterase; cholinesterase homology  
C:Keywords: carboxylic ester hydrolase  
F:24-541/Domain: cholinesterase homology <CHE>

Query Match 37.9%; Score 1086; DB 2; Length 544;  
Best Local Similarity 43.1%; Pred. No. 1.8e-70;  
Matches 236; Conservative 73; Mismatches 194; Indels 44; Gaps 9;

QY 15 PTAKLANGDTITGLNAIINEAFGLGIPFAEPVGNLRFKDPVPYSGSLNGQKFTSYGPGSCM 74  
DB 3 PTAVLNGNEISGLGVKVDTFKGIIPADPPVGLDRFKHPQFTGSGYQGLKANDFSACM 62  
QY 75 QONPEGTPE-----ENLGKLTALDLVMQSKVFAVLPOSEDCLTINVRPPGT 121  
DB 63 QLDPGNAISLWLDKVVGLGKILPDNLRGPLYDMA-----QGTVMSEDECLINVRPAGT 116  
QY 122 KAGANLPVWLMTFGGFGPIGSGPTIPPPAQMVTKSLVLMGKHIHVAVNYRVASWGLAGDD 181  
DB 117 KPDAKLPVWMIYGGAFVFGSSASYPGNGYVKESVEMGQPVVFSINRTGPGYFLGGDA 176  
QY 192 IKAGSGNAGLKQDLQGLGMQVADNIAGFGDPSKVTIFGESAGSMVSLCHLIWNGDNTY 241  
DB 177 ITAEGNTNAGLHDKQGLGWSDNIANFGGDPDKVMIIFGESAGSMVAHQIYVAYGDDNTY 236  
QY 242 KGKPLFRAGIMQSGAMVPSDPVDGTGYGNEIYDLFVSSAGC--SASDKLACLRASDDT 298  
DB 237 NGKKLFHSAIILQSGGLPYFDSTSVGPESASRFAQYAGCDTSDNDTLACLRKSSDV 296  
QY 299 LLDATNN-----TPGFLAYSSLRLSYLPRPGKNITDDMYKLVRDGVKVASVPVITG 349  
DB 297 LHSAQNSYDLKDLGLLQFLGFG-----PRPDGNIIPDAAYELYSRGYAKVPYITG 349  
QY 350 DONDEGTIFGLSNLNTTNAQARAYFKOSFIHASDAEITDMAAYPQDITOGSPDITGV 409  
DB 350 NQDEGTILAPVAINATTPHVKWLVKICSEASDASLDRLSLYPSGSEGAERTGIL 409  
QY 410 NALTPOPKRISAVLGLDLAFIHARRYFLNHFGOGTKYSPLSKQLSGL--PIMGTFFHNDIVW 468  
DB 410 NALTPOPKRIAAITDILLFQSPRRVNLNATKDVNRTYLATQLHNLVFLGTFFHGSDDL 469  
QY 469 QDYL-LGSGSVIYNNAFIATPDTLDPNTAGLLVNWPKYTSQSGNNLMNALGLYTGK 527  
DB 470 QYYVDLGPSSA--YRRYFISFANHHDPNVGTLNQWDMYT---DAGKEMLIQHMVGNMRT 525  
QY 528 DNERTAG 534  
DB 526 DDFRIEG 532

RESULT 12  
S41092  
triacylglycerol lipase (EC 3.1.1.3) I precursor - yeast (Geotrichum candidum) (strain CBS 178.71)  
C:Species: Geotrichum candidum  
A:Variety: NRRL Y-552  
C:Date: 19-Mar-1997 #sequence\_revision 05-Feb-1999 #text\_change 18-Jun-1999  
C:Accession: S41092  
R:Bertolini, M.C.; Laramee, L.; Thomas, D.Y.; Cygler, M.; Schrag, J.D.; Vernet, T.  
Eur. J. Biochem. 219, 119-125, 1994  
A:Title: Polymorphism in the lipase genes of Geotrichum candidum strains.  
A:Reference number: S41090; MUID:94139683; PMID:8306978  
A:Accession: S41092  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: DNA



A;Title: Molecular characterization of a Galactomyces geotrichum lipase, another member  
A;Reference number: S59957; MUID:96049515; PMID:7578238

A;Accession: S59957

A;Molecule type: nucleic acid

A;Residues: 1-563 <PH>

A;Cross-references: EMBL:X78032; NID:9459249; PIDN:CAA54963.1; PID:9459250

C;Superfamily: cholinesterase; cholinesterase homology

C;Keywords: carboxylic ester hydrolase

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-563/Product: triacylglycerol lipase #status predicted <MAT>

F;43-560/Domain: cholinesterase homology <CH>

Query Match 37.7%; Score 1081; DB 2; Length 563;  
Best Local Similarity 43.7%; Pred. No. 4.4e-70;  
Matches 240; Conservative 68; Mismatches 199; Indels 42; Gaps 8;

Qy 9 GRLGSVPTAKLANGDTITGLNAINEAPLGTPFAEPVGNLRFKDPVPYSGSLNGQKRTS 68

Db 16 GALAQTPFAVLNVNNOVISGLVGKVDTFKGIFFGAPPVGDRLRFKHPQPFQFTGYQGLKAND 75

Qy 69 YGSPCMQONPEGTF-----EENLGKTDALDVMQSKVFOAVLPOSEDCLTINV 115

Db 76 FSSACMLNPNALNTILDNALSLASISPENIRGLYDMAKGS-----VMSSEDCLYLVN 129

Qy 116 VRPGTKAGANLPMVLMTFGGFGIGSPITPPAQMTKSVLMGKHIIHVAVNRYVASWG 175

Db 130 CRPAGTKPGDKLPVMWMIYGAFTFGSSRSPGNGYVDESVKMGQPVVVFVSINVRSGPYG 189

Qy 176 FLAGDDIKAECSGNAGLKQORLGMQWADNITAGGDPSPKVTIFGESAGSMVLCILWN 235

Db 190 FLGGDGTAEAGNTNAGLHDQKGLWVSDNIANFGGDPCKVMIIFGESAGSMVGHLLTAY 249

Qy 236 DGDNTYKGPFLFRAGIMQSGAMVPSDPVDGTGYGNEIYDLFVSSAGC--SASDK--LACL 292

Db 250 GGDNTYNGKALFHSAILQSGPLPYNSGWLGPDSAYNRFAYAGCDTSASDVLEQLCLR 309

Qy 293 SASSDTLIDATNN-----TPGFLAYSSRLSYLPRPGKNITDDMYKLVLDGKYAS 343

Db 310 SKPSTSLDAQNSYDLKDLFGLLPQFLGFG-----PRPDGDIIPDSAYELYSRGYAK 362

Qy 344 VPVIIGQNDGTFIPLSSLVNTTNAQARAYFKQSFHASDAEIDTLMAAYPQDITQSP 403

Db 363 VPYITGQDEGTILAPVAINATTPHVKKWLKIFNEATDSLRLVILKLPETLSEGP 422

Qy 404 FDTGLNALTPQFKRIISAVLGDLAFIHARRYFLNHFQGGTKYFSLSKQLSGL--PIMGTFH 462

Db 423 FRTGLNALTPQFKRVAIITDILFQSPRDVMLNATKDVNWTFLATQLHNLVFFLGTFH 482

Qy 463 ANDIVQDYLGGSGSVIYNNAFATDLDNTAGLLVNNPKYTSSSQSGNNLMINALG 522

Db 483 ASDVLFQYLYNLIGPSDLYRFISFRNHDPNVGTGLQWAKYT---DGGKEMLEIRMLG 539

Qy 523 LYTCKDNFR 531

Db 540 NSMRTDDFR 548

RESULT 15

PN0493

Triacylglycerol lipase (EC 3.1.1.3) II precursor - yeast (Geotrichum candidum)

C;Species: Geotrichum candidum

C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 16-Jun-2000

C;Accession: PN0493; A46760

R;Nagao, T.; Shimada, Y.; Sugihara, A.; Tominaga, Y.

J. Biochem. 113, 776-780, 1993

A;Title: Cloning and sequencing of two chromosomal lipase genes from Geotrichum candidum

A;Reference number: PN0492; MUID:93380907; PMID:8370674

A;Accession: PN0493

A;Molecule type: DNA

A;Residues: 1-563 <NAG>

A;Note: the translation of residues 31-550 and the corresponding nucleotide sequence are

R;Shimada, Y.; Sugihara, A.; Iizumi, T.; Tominaga, Y.

J. Biochem. 107, 703-707, 1990

A;Title: cDNA cloning and characterization of Geotrichum candidum lipase II.  
A;Reference number: A46760; MUID:90375435; PMID:2398037

A;Accession: A46760

A;Molecule type: mRNA

A;Residues: 7-563 <SHI>

A;Cross-references: GB:D00697; NID:g217926; PIDN:BAA00603.1; PID:g217927

C;Comment: The extracellular lipase produced by Geotrichum candidum hydrolyzes all este

C;Genetics:

A;Gene: lipII

C;Superfamily: cholinesterase; cholinesterase homology

C;Keywords: carboxylic ester hydrolase; glycoprotein; pyroglutamic acid

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-563/Product: triacylglycerol lipase II #status predicted <MAT>

F;43-560/Domain: cholinesterase homology <CH>

F;234-238/Region: interfacial lipid recognition (GXSGXG) motif

F;20/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F;80-124,235-307/Disulfide bonds: #status predicted

F;236/Active site: Ser #status predicted

F;302,383/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 37.6%; Score 1078; DB 1; Length 563;

Best Local Similarity 41.6%; Pred. No. 7.2e-70;

Matches 233; Conservative 83; Mismatches 202; Indels 42; Gaps 8;

Qy 8 AGRILGSVPTAKLANGDTITGLNAINEAPLGTPFAEPVGNLRFKDPVPYSGSLNGQKFT 67

Db 15 AGVLAQAPRPSLNGNEVISGLVGKVDTFKGIFFADPPLDLRFKHPQPFQFTGSYQGLKAN 74

Qy 68 SYGSPCMQONPEGTFEENLGKTDALDVM--QSKVF-----QAVLPOSEDCLTIN 114

Db 75 DFSACMLNPNALNTILDNALSLASISPENIRGLYDMAKGTVMSEDCLYLVN 128

Qy 115 VVRPGTKAGANLPMVLMTFGGFGIGSPITPPAQMTKSVLMGKHIIHVAVNRYVASW 174

Db 129 VFRPAGTKPAKLPVMWMIYGAFTFGSSRSPGNGYVDESVKMGQPVVVFVSINVRGTF 188

Qy 175 GFLLAGDDIKAECSGNAGLKQORLGMQWADNITAGGDPSPKVTIFGESAGSMVLCILW 234

Db 189 GFLGGDAITAEAGNTNAGLHDQKGLWVSDNIANFGGDPDKVMIIFGESAGSMVAHQLIA 248

Qy 235 NDGNTYKGPFLFRAGIMQSGAMVPSDPVDGTGYGNEIYDLFVSSAGC---GSASDKLACL 291

Db 249 YGGDNTYNGKALFHSAILQSGPLPYNSGWLGPDSAYNRFAYAGCDTSASDVLECL 308

Qy 292 RSASSDTLLIDATNN-----TPGFLAYSSRLSYLPRPGKNITDDMYKLVLDGKYA 342

Db 309 RSKSSSVLHDAQNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELFRSGRYA 361

Qy 343 SVPVIIGQNDGTFIPLSSLVNTTNAQARAYFKQSFHASDAEIDTLMAAYPQDITQGS 402

Db 362 KVPYISGNQDEGTAFAPVALNATTPHVKKWLQYIFYDASEASIDRVLSLYPQTLVSGS 421

Qy 403 PFDTGVLNALTPOFKRISAVLGDLAFIHARRYFLNHFQGGTKYFSLSKQLSGL--PIMGTF 461

Db 422 PFTGILNALTPOFKRVAIISDMLFQSPRDVMLNATKDVNWTFLATQLHNLVFFLGTF 481

Qy 462 HANDIVWQDYLGGSGSVIYNNAFATDLDNTAGLLVNNPKYTSSSQSGNNLMINAL 521

Db 482 HGNELIFQFNVNIGPANSYLYRFISFANHDPNVGTNLLQWDQYT---DEGKEMLEIHTM 538

Qy 522 GLYTKDNFRTAGYDALMTN 541

Db 539 DNVWRTDDYRIEGISNPFETD 558

Search completed: July 29, 2004, 10:51:00

Job time : 23 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 10:50:34 ; Search time 50 Seconds  
(without alignments)

3431.691 Million cell updates/sec

Title: US-09-943-857-4

Perfect score: 2864

Sequence: 1 SMNSRGPAGRLGVSPTAKLA.....DNFRTAGYDALMTNPSFFV 547

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2864	100.0	547	10	US-09-943-857-4
2	2588	90.4	547	10	US-09-943-857-10
3	2495	87.1	547	10	US-09-943-857-8
4	2434	85.0	547	10	US-09-943-857-6
5	2396	83.7	547	10	US-09-943-857-2
6	1170.5	40.9	561	15	US-10-369-493-4236
7	674.5	23.6	612	9	US-09-875-353-4
8	672.5	23.5	574	13	US-10-023-515-4
9	672.5	23.5	574	16	US-10-674-636-4
10	664	23.2	585	9	US-09-934-323-4
11	482	16.8	574	9	US-09-748-739A-21
12	482	16.8	574	14	US-10-032-233-48
13	482	16.8	574	16	US-10-324-466-48
14	482	16.8	574	16	US-10-413-432-48
15	477	16.7	574	9	US-09-748-739A-23

16	477	16.7	574	14	US-10-032-233-50	Sequence 50, Appl
17	477	16.7	574	16	US-10-324-466-50	Sequence 50, Appl
18	477	16.7	574	16	US-10-413-432-50	Sequence 50, Appl
19	459	16.0	574	9	US-09-748-739A-22	Sequence 22, Appl
20	459	16.0	574	14	US-10-032-233-49	Sequence 49, Appl
21	459	16.0	574	16	US-10-324-466-49	Sequence 49, Appl
22	459	16.0	574	16	US-10-413-432-49	Sequence 49, Appl
23	456	15.9	574	14	US-10-032-233-40	Sequence 40, Appl
24	456	15.9	574	16	US-10-324-466-40	Sequence 40, Appl
25	456	15.9	574	16	US-10-413-432-40	Sequence 40, Appl
26	455	15.9	574	9	US-09-748-739A-8	Sequence 8, Appl
27	452	15.8	574	14	US-10-032-233-34	Sequence 34, Appl
28	452	15.8	574	16	US-10-324-466-34	Sequence 34, Appl
29	452	15.8	574	16	US-10-413-432-34	Sequence 34, Appl
30	452	15.8	614	15	US-10-116-275-258	Sequence 258, App
31	451	15.7	574	9	US-09-748-739A-6	Sequence 6, Appl
32	451	15.7	574	9	US-09-748-739A-17	Sequence 17, Appl
33	451	15.7	574	9	US-09-748-739A-20	Sequence 20, Appl
34	451	15.7	574	10	US-09-997-209-89	Sequence 89, Appl
35	451	15.7	574	14	US-10-032-233-22	Sequence 22, Appl
36	451	15.7	574	14	US-10-032-233-24	Sequence 24, Appl
37	451	15.7	574	14	US-10-032-233-26	Sequence 26, Appl
38	451	15.7	574	14	US-10-032-233-44	Sequence 44, Appl
39	451	15.7	574	14	US-10-032-233-47	Sequence 47, Appl
40	451	15.7	574	15	US-10-326-892-2	Sequence 2, Appl
41	451	15.7	574	16	US-10-433-206-89	Sequence 89, Appl
42	451	15.7	574	16	US-10-324-466-22	Sequence 22, Appl
43	451	15.7	574	16	US-10-324-466-24	Sequence 24, Appl
44	451	15.7	574	16	US-10-324-466-26	Sequence 26, Appl
45	451	15.7	574	16	US-10-324-466-44	Sequence 44, Appl

#### ALIGNMENTS

RESULT 1  
US-09-943-857-4  
; Sequence 4, Application US/09943857  
; Publication No. US20030124701A1  
; GENERAL INFORMATION:  
; APPLICANT: Shaw, Jei-Fu  
; APPLICANT: Lee, Guan-Chiun  
; APPLICANT: Tang, Shye-Jye  
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES  
; FILE REFERENCE: 08919-066001  
; CURRENT APPLICATION NUMBER: US/09/943,857  
; CURRENT FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 547  
; TYPE: PRT  
; ORGANISM: Candida rugosa  
; US-09-943-857-4

Query Match	100.0%;	Score 2864;	DB 10;	Length 547;
Best Local Similarity	100.0%;	Pred. No. 2.8e-265;		
Matches 547;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	SMNSRGPAGRLGVSPTAKLANGDTITGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGS	60	
Db	1	SMNSRGPAGRLGVSPTAKLANGDTITGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGS	60	
Qy	61	LNQCKTSTYGPSQCNQPECTFEENLGKTDALVMSKVQAVLPQSEDCLTINVRPPG	120	
Db	61	LNQCKTSTYGPSQCNQPECTFEENLGKTDALVMSKVQAVLPQSEDCLTINVRPPG	120	
Qy	121	TKAGANLPVWLWTFGGGFEIGSPITIPPAOMVTKSVLMGKHIIHVAIVYVAVSWGFLAGD	180	
Db	121	TKAGANLPVWLWTFGGGFEIGSPITIPPAOMVTKSVLMGKHIIHVAIVYVAVSWGFLAGD	180	
Qy	181	DIKAEGSGNAGLQDKRIGMQWADNIAGFGDPSKVTIFGESAGSMVLCHELLINWDGNT	240	
Db	181	DIKAEGSGNAGLQDKRIGMQWADNIAGFGDPSKVTIFGESAGSMVLCHELLINWDGNT	240	



Db 181 DIKAGSGNAGLKQORLGMQWADNIAGFGDPSKVTIFGESAGSMSVLCHLIWNGDNT 240  
Qy 241 YKGKPLFRAGIMQSGAMVPSDPVDTGTYGNEIYDLFVSSAGCGSADKLACLRSSDITLL 300  
Db 241 YKGKPLFRAGIMQSGAMVPSDPVDTGTYGNEIYDLFVSSAGCGSADKLACLRSSDITLL 300  
Qy 301 DATNNTPGFLAYSSRLSRLSYLPRPDGKNTTDDMYKLVRDGYKASVPVVIIGDQNDGRTIFGL 360  
Db 301 DATNNTPGFLAYSSRLSRLSYLPRPDGKNTTDDMYKLVRDGYKASVPVVIIGDQNDGRTIFGL 360  
Qy 361 SSLNVTNNAQARAYFKQSFIIHASDAEIDTLMAAYPQDITQSGPDTGVNALTPOFKRIS 420  
Db 361 SSLNVTNNAQARAYFKQSFIIHASDAEIDTLMAAYPQDITQSGPDTGVNALTPOFKRIS 420  
Qy 421 AVLDGLAFIHARRYFLNHFGQGTKYSLKQSLGPIVGTGTHANDIWMQDYLGGSGVIY 480  
Db 421 AVLDGLAFIHARRYFLNHFGQGTKYSLKQSLGPIVGTGTHANDIWMQDYLGGSGVIY 480  
Qy 481 NNATIAFATDLPNTAGLLVNWPKYTSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540  
Db 481 NNATIAFATDLPNTAGLLVNWPKYTSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540  
Qy 541 NPSSFFV 547  
Db 541 NPSSFFV 547

RESULT 2  
US-09-943-857-10  
; Sequence 10, Application US/09943857  
; Publication No. US20030124701A1  
; GENERAL INFORMATION:  
; APPLICANT: Shaw, Jai-Fu  
; APPLICANT: Lee, Guan-Chiun  
; APPLICANT: Tang, Shye-Jye  
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES  
; FILE REFERENCE: 08919-066001  
; CURRENT APPLICATION NUMBER: US/09/943.857  
; CURRENT FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 547  
; TYPE: PRT  
; ORGANISM: Candida rugosa  
US-09-943-857-10

Query Match 90.4%; Score 2588; DB 10; Length 547;  
Best Local Similarity 88.7%; Pred. No. 8.1e-239;  
Matches 485; Conservative 30; Mismatches 32; Indels 0; Gaps 0;

Qy 1 SMNSRGPAGRLGVSPTAKLANGDTITGLNAINAEFLGIPFAEPPVGNLRKDPVPYSGS 60  
Db 1 SMNSRGPAGRLGVSPTAKLANGDTITGLNAINAEFLGIPFAEPPVGNLRKDPVPYSGS 60

Qy 61 LKGKFTSYGSPSCMQQNEGTFFENLGTALDLVMQSKVFOAVLPQSEDCITINVRPPG 120  
Db 61 LKGKFTSYGSPSCMQQNEGTFFENLGTALDLVMQSKVFOAVLPQSEDCITINVRPPG 120

Qy 121 TKAGANLPVMLWIFGGGFEIGSPFIAPPQWVTKSVMGMKHIHVVANNYRVASWGFLAGD 180  
Db 121 TKAGANLPVMLWIFGGGFEIGSPFIAPPQWVTKSVMGMKHIHVVANNYRVASWGFLAGD 180

Qy 181 DIKAGSGNAGLKQORLGMQWADNIAGFGDPSKVTIFGESAGSMSVLCHLIWNGDNT 240  
Db 181 DIKAGSGNAGLKQORLGMQWADNIAGFGDPSKVTIFGESAGSMSVLCHLIWNGDNT 240

Qy 241 YKGKPLFRAGIMQSGAMVPSDPVDTGTYGNEIYDLFVSSAGCGSADKLACLRSSDITLL 300  
Db 241 YKGKPLFRAGIMQSGAMVPSDPVDTGTYGNEIYDLFVSSAGCGSADKLACLRSSDITLL 300

Qy 301 DATNNTPGFLAYSSRLSRLSYLPRPDGKNTTDDMYKLVRDGYKASVPVVIIGDQNDGRTIFGL 360  
Db 301 DATNNTPGFLAYSSRLSRLSYLPRPDGKNTTDDMYKLVRDGYKASVPVVIIGDQNDGRTIFGL 360

Db 301 DATNNTPGFLAYSSRLSRLSYLPRPDGKNTTDDMYKLVRDGYKASVPVVIIGDQNDGRTIFGL 360  
Qy 361 SSLNVTNNAQARAYFKQSFIIHASDAEIDTLMAAYPQDITQSGPDTGVNALTPOFKRIS 420  
Db 361 SSLNVTNNAQARAYFKQSFIIHASDAEIDTLMAAYPQDITQSGPDTGVNALTPOFKRIS 420  
Qy 421 AVLDGLAFIHARRYFLNHFGQGTKYSLKQSLGPIVGTGTHANDIWMQDYLGGSGVIY 480  
Db 421 AVLDGLAFIHARRYFLNHFGQGTKYSLKQSLGPIVGTGTHANDIWMQDYLGGSGVIY 480  
Qy 481 NNATIAFATDLPNTAGLLVNWPKYTSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540  
Db 481 NNATIAFATDLPNTAGLLVNWPKYTSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540  
Qy 541 NPSSFFV 547  
Db 541 NPSSFFV 547

RESULT 3  
US-09-943-857-8  
; Sequence 8, Application US/09943857  
; Publication No. US20030124701A1  
; GENERAL INFORMATION:  
; APPLICANT: Shaw, Jai-Fu  
; APPLICANT: Lee, Guan-Chiun  
; APPLICANT: Tang, Shye-Jye  
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES  
; FILE REFERENCE: 08919-066001  
; CURRENT APPLICATION NUMBER: US/09/943.857  
; CURRENT FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 547  
; TYPE: PRT  
; ORGANISM: Candida rugosa  
US-09-943-857-8

Query Match 87.1%; Score 2495; DB 10; Length 547;  
Best Local Similarity 86.3%; Pred. No. 6.7e-230;  
Matches 472; Conservative 31; Mismatches 44; Indels 0; Gaps 0;

Qy 1 SMNSRGPAGRLGVSPTAKLANGDTITGLNAINAEFLGIPFAEPPVGNLRKDPVPYSGS 60  
Db 1 SMNSRGPAGRLGVSPTAKLANGDTITGLNAINAEFLGIPFAEPPVGNLRKDPVPYSGS 60

Qy 61 LKGKFTSYGSPSCMQQNEGTFFENLGTALDLVMQSKVFOAVLPQSEDCITINVRPPG 120  
Db 61 LKGKFTSYGSPSCMQQNEGTFFENLGTALDLVMQSKVFOAVLPQSEDCITINVRPPG 120

Qy 121 TKAGANLPVMLWIFGGGFEIGSPFIAPPQWVTKSVMGMKHIHVVANNYRVASWGFLAGD 180  
Db 121 TKAGANLPVMLWIFGGGFEIGSPFIAPPQWVTKSVMGMKHIHVVANNYRVASWGFLAGD 180

Qy 181 DIKAGSGNAGLKQORLGMQWADNIAGFGDPSKVTIFGESAGSMSVLCHLIWNGDNT 240  
Db 181 DIKAGSGNAGLKQORLGMQWADNIAGFGDPSKVTIFGESAGSMSVLCHLIWNGDNT 240

Qy 241 YKGKPLFRAGIMQSGAMVPSDPVDTGTYGNEIYDLFVSSAGCGSADKLACLRSSDITLL 300  
Db 241 YKGKPLFRAGIMQSGAMVPSDPVDTGTYGNEIYDLFVSSAGCGSADKLACLRSSDITLL 300

Qy 301 DATNNTPGFLAYSSRLSRLSYLPRPDGKNTTDDMYKLVRDGYKASVPVVIIGDQNDGRTIFGL 360  
Db 301 DATNNTPGFLAYSSRLSRLSYLPRPDGKNTTDDMYKLVRDGYKASVPVVIIGDQNDGRTIFGL 360

Qy 361 SSLNVTNNAQARAYFKQSFIIHASDAEIDTLMAAYPQDITQSGPDTGVNALTPOFKRIS 420  
Db 361 SSLNVTNNAQARAYFKQSFIIHASDAEIDTLMAAYPQDITQSGPDTGVNALTPOFKRIS 420

Qy 421 AVLDGLAFIHARRYFLNHFGQGTKYSLKQSLGPIVGTGTHANDIWMQDYLGGSGVIY 480  
Db 421 AVLDGLAFIHARRYFLNHFGQGTKYSLKQSLGPIVGTGTHANDIWMQDYLGGSGVIY 480



Db 421 AVLGLDFTLTSRRYFLNHVYTGPKYFSLSKQLSGLPILGTGFHANDIVWQHFLGSGSVIY 480  
QY 481 NNAFTAFATDLPNTAGLLVNMKPTSSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540  
Db 481 NNAFTAFATDLPNTAGLSVQMPKPTSSSQAGDNLMOISALGLYTGKDNFRTAGYNALFA 540  
QY 541 NPSSEFFV 547  
Db 541 DPSHFFV 547

RESULT 4  
US-09-943-857-6  
; Sequence 6, Application US/09943857  
; Publication No. US20030124701A1  
; GENERAL INFORMATION:  
; APPLICANT: Shaw, Jei-Fu  
; APPLICANT: Lee, Guan-Chiun  
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES  
; FILE REFERENCE: 08919-066001  
; CURRENT APPLICATION NUMBER: US/09/943,857  
; CURRENT FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 547  
; TYPE: PRT  
; ORGANISM: Candida rugosa  
US-09-943-857-6

Query Match 85.0%; Score 2434; DB 10; Length 547;  
Best Local Similarity 83.5%; Pred. No. 4.7e-224;  
Matches 457; Conservative 38; Mismatches 52; Indels 0; Gaps 0;

QY 1 SMNSRGPAGRLGVSPTAKLANGDTITGLNAINAEFLGIPPAEPVGNLRFKDPVPYSGS 60  
Db 1 SMNSRGPAGRLGVSPTATLANGDTITGLNAINAEFLGIPPAEPVGNLRFKDPVPYSGS 60  
QY 61 LINGQFTSYGSCMQNPEGTPEENLKGKLTALDLVMSKVQAVLPQSEDECLINVRPPG 120  
Db 61 LINGQFTSYGSCMQNPLGNDSSLPKAAINSIMQSKLFQAVLPNGEDCLINVRPPG 120  
QY 121 TKAGANLPMWLIFGGGFEIGSPTIPPPAQMTKSVLMGKHIIHVAVNVRVASWGLAGD 180  
Db 121 TKPGANLPMWLIFGGGFEVGGSSLPFAQMITASVLMGKPIIHVSMYRVRASWGLAGP 180  
QY 181 DIKAGSGNAGLKDORLGMQWADNIAGFGDPKVTIFGESAGSMVLCGLIWNDDGNT 240  
Db 181 DIKAGSGNAGLHDQRGLQWADNIAGFGDPKVTIFGESAGSMVLMCQLLWDDGNT 240  
QY 241 YGKELFRAGIMQSGAMVPSDPVDTYGNIEYDLFVSSAGCGSADKLCLRSASDITLL 300  
Db 241 YNGKELFRAMQSGAMVPSDPVDTYGNIEYDVVVASAGCGSADKLCLRSISNDKLF 300  
QY 301 DATNTPGLAYSLRLSRLSPDPGKNITDDMYKLVROGKYASVPVIGQDNDEGTIFGL 360  
Db 301 QATSTPGALAYPSRLSRLSPDPGDTITDDMFKLVROGKCANVPVIGQDNDEGTIFAL 360  
QY 361 SSLNVTNQAARAYFKQSFHASDAEIDTLMAAYPQDITQGSPPFTGVLNALTPOPKRIS 420  
Db 361 SSLNVTDAQARQYFKESFIHASDAEIDTLMAAYPSDITQGSPPFTGTFNALTPOPKRIA 420  
QY 421 AVLGDLAFTHARRYFLNHFFQGGTKYFSLSKQLSGLPILGTGFHANDIVWQDYLGGSVIY 480  
Db 421 AVLGDLAFTLPRYFLNHFFQGGTKYFSLSKQLSGLPILGTGFHANDIVWQDYLSSHSAVY 480  
QY 481 NNAFTAFATDLPNTAGLLVNMKPTSSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540  
Db 481 NNAFTAFANDLPNKAGLLVNMKPTSSSQSGNNLQINALGLYTGKDNFRTAGYDALMT 540  
QY 541 NPSSEFFV 547  
Db 541 DPSHFFV 547

Db 541 NPSSEFFV 547

RESULT 5  
US-09-943-857-2  
; Sequence 2, Application US/09943857  
; Publication No. US20030124701A1  
; GENERAL INFORMATION:  
; APPLICANT: Shaw, Jei-Fu  
; APPLICANT: Lee, Guan-Chiun  
; APPLICANT: Tang, Shye-Jye  
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES  
; FILE REFERENCE: 08919-066001  
; CURRENT APPLICATION NUMBER: US/09/943,857  
; CURRENT FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 547  
; TYPE: PRT  
; ORGANISM: Candida rugosa  
US-09-943-857-2

Query Match 83.7%; Score 2396; DB 10; Length 547;  
Best Local Similarity 82.3%; Pred. No. 2.1e-220;  
Matches 450; Conservative 34; Mismatches 63; Indels 0; Gaps 0;

QY 1 SMNSRGPAGRLGVSPTAKLANGDTITGLNAINAEFLGIPPAEPVGNLRFKDPVPYSGS 60  
Db 1 SMNSRGPAGRLGVSPTATLANGDTITGLNAINAEFLGIPPAEPVGNLRFKDPVPYSGS 60  
QY 61 LINGQFTSYGSCMQNPEGTPEENLKGKLTALDLVMSKVQAVLPQSEDECLINVRPPG 120  
Db 61 LINGQFTSYGSCMQNPMGSPEDTLPKNALDLVLSKLFQVVLNDEDECLINVRPPG 120  
QY 121 TKAGANLPMWLIFGGGFEIGSPTIPPPAQMTKSVLMGKHIIHVAVNVRVASWGLAGD 180  
Db 121 TRASAGLPMWLIFGGGFEIGSSLPFGQWVAKSVLMGKPIIHVSMYRVRASWGLAGP 180  
QY 181 DIKAGSGNAGLKDORLGMQWADNIAGFGDPKVTIFGESAGSMVLCGLIWNDDGNT 240  
Db 181 DIQNGSGNAGLHDQRGLQWADNIAGFGDPKVTIYGESAGSMSTFVHLVWDDGNT 240  
QY 241 YGKELFRAGIMQSGAMVPSDPVDTYGNIEYDLFVSSAGCGSADKLCLRSASDITLL 300  
Db 241 YNGKELFRAMQSGAMVPSDPVDTYGNIEYQVVASAGCGSADKLCLRLSQTLY 300  
QY 301 DATNTPGLAYSLRLSRLSPDPGKNITDDMYKLVROGKYASVPVIGQDNDEGTIFGL 360  
Db 301 QATSTPGVLAYPELSRLSRLSPDPGDTITDDMYALVRDGYAHVPVIGQDNDEGTIFGL 360  
QY 361 SSLNVTNQAARAYFKQSFHASDAEIDTLMAAYPQDITQGSPPFTGVLNALTPOPKRIS 420  
Db 361 SSLNVTDAQARAYFKQSFHASDAEIDTLMAAYTSDITQGSPPFTGTFNALTPOPKRIS 420  
QY 421 AVLGDLAFTHARRYFLNHFFQGGTKYFSLSKQLSGLPILGTGFHANDIVWQDYLGGSVIY 480  
Db 421 ALLGDLAFTLARRYFLNLYYQGGTKYFSLSKQLSGLPVLGTGHGNDIIMQDYLVGSGSVIY 480  
QY 481 NNAFTAFATDLPNTAGLLVNMKPTSSSQSGNNLMINALGLYTGKDNFRTAGYDALMT 540  
Db 481 NNAFTAFANDLPNKAGLLVNMKPTSSSQSGNNLQINALGLYTGKDNFRPRDAYSAIFS 540  
QY 541 NPSSEFFV 547  
Db 541 NPSSEFFV 547

RESULT 6  
US-10-369-493-4236  
; Sequence 4236, Application US/10369493  
; Publication No. US2003023675A1  
; GENERAL INFORMATION:

```

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 4236
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Neurospora crassa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(561)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-4236

Query Match      40.9%; Score 1170.5; DB 15; Length 561;
Best Local Similarity 43.6%; Pred. No. 7.2e-103;
Matches 247; Conservative 96; Mismatches 175; Indels 49; Gaps 10;

Qy 13 SVPTAKLANGDTITGLNAINEAEFLGIPPAEPVGNLRPKDPVPPYSGSLNGQKFTSYGPS 72
Db 1 AAPTITSTG-TIVGANGILTEAFNGIPYALPPTGNLRKPPVRLKSSLVDFDASGIGPA 59
Qy 73 CMOQNPGTEFEENLGKLTALDVMQSKVFOAVLPQSEDECLTINVVRPPTKAGANLPVLMW 132
Db 60 CPQFLATSSNEFLPQ-VIDKIVNTQFLKILNVKEDCLISVTRPKTKAGDKLPVLF 118
Qy 133 IFGGGFPI-----GSPTIFPPAQMTKSVLMGKHIIHVAVNRYVASW 174
Db 119 IFGGGFVVRNPAYTRDVMNCKNFTAGSASMYDGAPLVTNAINMGKPYVAVNRYVGGF 178
Qy 175 GFLAGDDIKAEKSGNAGLKQORLCQWADNIAGFGGDPKVTIFGESAGSMVLCILW 234
Db 179 GFMPGKBILKDGSSNLGHLDDQWGLQWADNIAAFGGDDPKVTITWGSAGSMVFNQMSL 238
Qy 235 NKGNTYKGLPLFRAGIMQSGAMVPDPVDCGTGNEIYDLFVSSAGSGSASDKLCLRSA 294
Db 239 YDGNNTYNGKPLFRGAIMWSSGIYPAGVDCPKQKYDITVKNAGCGSADTLCLRAL 298
Qy 295 SSDTLDATNNTPGFLAYSSRLSLRPLRPGKNTTDDMYKLVRDGYKVASVPVIGDQND 354
Db 299 PYETFLKAANSVPGLSYNSVALSPLRPGDKALTQSADKIMLAKKYAAVPMIIGDQDE 358
Qy 355 GTIFGLSLLVNTTNAQARAYPKQSIH-ASDAEIDTLMAAYPODITOGSPDTCVLNALT 413
Db 359 GTLFSLFQSNITTSKLVSYNDIFFNDATESQIKLSVYSTLTISAGSPFGTLFNEYI 418
Qy 414 PQFKRISAVLGLDLAFIHARRYFLNHFQGGT-----KYSFLSKQLSGCLPTMGTFHANI 466
Db 419 PGFKRLAAILGLDILFTLSRIIFLD---AATLNPSPVAMSVLASYNFTGPILGTFHADI 475
Qy 467 --VWQDYLLGSGSVIYNNAIFATDLDPNAG-----LIVNPKYTSOSSGNNLM 516
Db 476 LOVFYGLPLPNYASKSIQSYVANFYNLDPNDASGTSKSKVQDMPWQOKERK----- 529
Qy 517 MINALGYLTC--KDNRTAGYDALMTN 541
Db 530 LVQFFSDYAGVLADDFRSDSYNWKAN 556

```

RESULT 7  
 US-09-875-353-4  
 ; Sequence 4, Application US/09875353  
 ; Patent No. US20020168713A1  
 ; GENERAL INFORMATION:

```

; APPLICANT: Curtis, Rory A.J.
; TITLE OF INVENTION: 46980, A NOVEL HUMAN NEUROLIGIN FAMILY
; TITLE OF INVENTION: MEMBER AND USES THEREOF
; FILE REFERENCE: 10448-058001
; CURRENT APPLICATION NUMBER: US/09/875,353
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/209,949
; PRIOR FILING DATE: 2000-06-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 612
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus sequence
US-09-875-353-4

Query Match      23.6%; Score 674.5; DB 9; Length 612;
Best Local Similarity 34.0%; Pred. No. 3e-55;
Matches 199; Conservative 80; Mismatches 203; Indels 103; Gaps 19;

Qy 19 LANGDITITGLNAINE-----AFLGIPPAEPVGNLRPKDPVPPY-----SGSLNGQ 64
Db 31 VATNNVLICGRVGRVNEKTDNGEQSVSYFLGIPYAEPPVGNLRPKAPQPYKEPWSVDVLD 89
Qy 65 KFTSYGSPCMQNPQEGTFEENLGKLTALDVMQSKVFOAVLPQSEDECLTINVVRPPTKAG 124
Db 90 --TKYPSCQLQDDDFGSLSDL-KVALKMLSLGNKLVGLKLSDECLYLVNYPKTKPN 146
Qy 125 ANLPMVLWIFGGGFEIGSPTIFPPAQMTKSVLMGKHIIHVAVNRYVASWGLAGDDIKA 184
Db 147 SKLPVMVWIHGGGFMFGSGHSLPLSLYDGSGLAREGNVIVVSYNYRLGLPGLSTGDDKL 206
Qy 185 EGSNAGLKQORLCQWADNIAGFGGDPKVTIFGESAGSMVLCILWINDGNTYK 244
Db 207 PGSGNYGLLDQRLALKWVQDNIAAFGGDPNSVTIFGESAGAAVSLLLLSNGGDNPPSSK 266
Qy 245 PLFRAGIMQSGA-----MVPSPDVGTYGNEIYDLFVSSAGCG--SASDKLACLRSASSD 297
Db 267 GLFHRAISQSGSALSFWAIQSESNARGAKELARLL---GCNETSSELDCLRKSAE 322
Qy 298 TLLDATNNTPGFLAYS-----SLRLSYLPRPDGKN-----ITDDMYKLVRDGYKVASVPVI 347
Db 323 ELLEATRS---FLLFEYVPLPLFLAPGVVDGDAPEAFIPEDPRELIKEGKPADVPYL 379
Qy 348 IGOQNBCTTIFGLSLLVNTTNAQAR-----AYFKQSFHASDA-----EID 388
Db 380 IGVTKDEGGYFAAMLNLSKSGEDELKKTNPDPVWLELLKYLIFYASEALNIKMDMDLAD 439
Qy 389 TLMAAYPODITOGSPDTCVLNALTPOFKRISAVLGLDLAF-----THARR----- 433
Db 440 KVLKPYDGVDDFS-----VESRKNLQDMLTDLFLKCPTRVADLHAKHGSGSPVY 490
Qy 434 -YFINHFQGGTKSYFLSKQLSGCLPTMGTFHANI--VWQDYLL-----GSGSV 478
Db 491 AYVFDHPASPFGIGQFLAKRVDP-EFGAVHGDDEIFFVFGNPLLKEQLYKATEBEKSSSK 549
Qy 479 IYNNAFATATDLDPN--TAGLLVNWPKYTSOSSGNNLMINAL 521
Db 550 TMNMYANFAKTGNPNNGTNSGLVWVWPKYTSSEQKYSLLILLTTI 594

RESULT 8
US-10-023-515-4
; Sequence 4, Application US/10023515
; Publication No. US20020182636A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Silos-Santiago, Immaculada
; TITLE OF INVENTION: 53010, A NOVEL HUMAN CARBOXYLESTERASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREOF
; FILE REFERENCE: 10448-122001

```

```

; CURRENT APPLICATION NUMBER: US/10/023,515
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 60/256,369
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/279,508
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
US-10-023-515-4

```

Query Match	23.5%	Score	672.5	DB 13	Length	574
Best Local Similarity	34.9%	Pred. No.	4.2e-55			
Matches 195	Conservative	76	Mismatches	195	Indels	93
					Gaps	18

RESULT 9  
US-10-674-636-4  
; Sequence 4, Application US/10674636  
; Publication No. US20040086922A1  
; GENERAL INFORMATION:  
; APPLICANT: Curtis, Rory A. J.  
; APPLICANT: Siles-Santiago, Inmaculada  
; TITLE OF INVENTION: 53010, A NOVEL  
; TITLE OF INVENTION: FAMILY MEMBER  
; FILE REFERENCE: 10448-122001  
; CURRENT APPLICATION NUMBER: US/10/  
; CURRENT FILING DATE: 2003-09-29

```

, PRIOR APPLICATION NUMBER: US/10/023,515
, PRIOR FILING DATE: 2001-12-18
, PRIOR APPLICATION NUMBER: 60/256,369
, PRIOR FILING DATE: 2000-12-18
, PRIOR APPLICATION NUMBER: 60/279,508
, PRIOR FILING DATE: 2001-03-28
, NUMBER OF SEQ ID NOS: 6
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 4
, LENGTH: 574
, TYPE: PRT
, ORGANISM: Artificial Sequence
, FEATURE:
, OTHER INFORMATION: Consensus sequence
US-10-674-636-4

```

Query Match.	23.5%;	Score 672.5;	DB 16;	Length 574;
Best Local Similarity	34.9%;	Pred. No. 4.2e-55;		
Matches 195;	Conservative	76;	Mismatches 195;	Indels 93; Gaps 18
Qy	35	AFLGIPAEPPVGNLRKDPVPY-----SGSLNGQKFTSYGPSCMQONPEGTFEENLGKTA 90		
Db	19	SFLGIPAEPPVGNLRKFAQPYKEPWSVLDA---TKYPSCIQDDDFGFSLSDL-KVA 74		
Qy	91	LDLVMQSKVFOAVLPQSEDCLTITINVPPGTKAGANLPVMLWIEFGGFEIGSPITFPFPAQ 150		
Db	75	LKMLSLGNWKLVLKLSDECLYLVNYPKNTKPSKLPVMVMIHGGFMFGSGHSLPLSL 134		
Qy	151	MVTKSVLMGXHIHVAVNRYVASWGFAGDDIKAEGSGNAGLKDOORLCMQWADNIAGFG 210		
Db	135	YDGESLAREGNVIWVSINRYRGPJGFJSTGDDKLPGSGNYGLLDQRLAKWQDNIAAF 194		
Qy	211	GDPKSVTIFGESAGSMVLCHLIWNDDGNTYKGPLFRAGIMQSGA----MVPSPDPVDG 265		
Db	195	GDPNSVTIFGESAGAAVSLLLLSNGDNPSSKGLFHRAISQGSALSPWAIQESNAR 254		
Qy	266	TYGNEIYDLFWSSAGCG--SASDKLACLRSSASTLLDATNNTPGFLAYS-----SLRLS 318		
Db	255	GRAKELARLL---GCNETSSSELDCIRKSABELLEATRS---FLLFYVVPFLPLFLA 307		
Qy	319	YLPRPDCKN-----ITDDMYKLVDGKVASVPVLIQNDQDGTGFLGSLNAVTTTNAQAR- 372		
Db	308	FGPVVDGDDAPEAFIPEDPBELIKEGFADVPYHIGVTKDEGGFPAAMLLNASSKGEDEL 367		
Qy	373	-----AYFKQSITHASDA-----EIDTLMAAPQDITQGSFDTGVLNALTP 414		
Db	368	KKETNPVWLELLKYLLEYASEALNIKMDDLADKLVLEKYPGDVDDFS-----VES 418		
Qy	415	QFKRISAVLGDIAF-----IHARR-----YFLNHFGQGTKYSFLSKOLSGLPIM 458		
Db	419	RKPNLOMLTDLPLKCPTRVAADLJHAKHGGSPVYAYVFDHPASFQIGOFQAKRVDP-BFG 477		
Qy	459	GTFHANDI--VWQDYLL-----GSGSVIYNNAFIATAFDLDPN--TAGLLVNW 502		
Db	478	GAVHGDIFFVFGPNLLKEQLYKATEBEERSSSKTMNYWANFAKTGNPNNGTSLGLVW 537		
Qy	503	PKYTSSSQSGNNLMINAL 521		
Db	538	PKYTSEOKYSLILLTTI 556		

RESULT 10  
US-09-934-323-4  
; Sequence 4, Application US/09934323  
; Patent No. US20020150910A1  
; GENERAL INFORMATION:  
; APPLICANT: Curtis, Rory A. J.  
; TITLE OF INVENTION: 33410. A NOVEL HUMAN CARBOXYLESTERASE  
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREOF  
; FILE REFERENCE: 10448-081001  
; CURRENT APPLICATION NUMBER: US/09/934,323  
; CURRENT FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: US 60/226,774

```
; PRIOR FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
US-09-934-323-4

Query Match      23.2%; Score 664; DB 9; Length 585;
Best Local Similarity 34.0%; Pred. No. 2.8e-54;
Matches 199; Conservative 80; Mismatches 203; Indels 104; Gaps 20;

Qy 19 LANGDTITGLNAIN-----AFLGIPFAEPVGNLRFKDPVPY-----SGSLNGQ 64
Db 3 VATNNVLGKVRGYNKTDNGEQSVYSLGIPYAEPPVGNLRFKAPQYKEPWSVDVLD- 61
Qy 65 KFTSYGSCMOONPEGTFEENLGTALDLVMSQVFOAVLPQSEDCLTINVRPPGKAG 124
Db 62 --TKYPSCLQDDFGFSLDL--KVALKMLSLGNLKLSEDCLYLVNVTYKYNKP 118
Qy 125 ANLPMVLWIFGGGFEIGSPITFFPAQMVTKSVLMKHHIVAVNRYRVASWGLAGDDIKA 184
Db 119 SKLPVWIIHGGFMFGSGHSLPLSLYDGSLEAREGNVIVSVINRLGPLGFLSTGDDKL 178
Qy 185 EGSGNAG-LKDQRLGMQWADNIAGFGDPKSVTIFGESAGSMVSLCHLIWNGDNTYKG 243
Db 179 PGSGNYGLLLDQRLALKWVQDNIAAFGGDPNSVTIFGESAGAAVSULLLSNGGDNPPSS 238
Qy 244 KPLFRAGIMQSGA----MVPSPDVGTYGNEIYDLFVSSAGCG--SASDKLACLRSASS 296
Db 239 KGLFRAISQSGSALSPWAIQSESNARGRAKELARLL----GCNETSSSELLDCLRSKA 294
Qy 297 DTLADATNNTPGFLAYS-----SLRLSYLPRPDGKN-----ITDDMYKLVRDGYKASVPV 346
Db 295 EELLEATRS---FLLFYVVPFLPLFLAPGPVVDGDAPEAFIPEDPEBELKEGKFADVPY 351
Qy 347 IIGQNDGTTPLGLSSLVNTTNAQAR-----AYFKQSFHASDA-----EI 387
Db 352 LIGVTKDEGGYFAAMLLNASSGDELLKKTETNPVWLELLKYLIFYASEALNIKMDMDLA 411
Qy 388 DTLMAAYPDITQSPDPTGVNLALTPOFKISAVLGLDAF-----IHARR-----433
Db 412 DKVLEKTPGDVDDFS-----VESRKNLQMDLTDLLPKCTPRVAADLHAKHGSPV 462
Qy 434 --YFLNHFQGGTKYSFSLKQSLGSLPIMGTTHANDI--VWQDYLL-----GSGS 477
Db 463 YAYVFDHPASFGIGQFLAKRVDP-EFGGAVHGDEIFFVFGNPLKEQLYKATEBEEKSSS 521
Qy 478 VIYNNAFTAFATDLPN--TAGLLVWMPKYTSOSSQSGNNLAMINAL 521
Db 522 KTMNYWANFAKTGNPNNGTNGLVVWMPKYTSBEEKYSLILLITTI 567

RESULT 11
US-09-748-739A-21
; Sequence 21, Application US/09748739A
; Patent No. US20020119489A1
; GENERAL INFORMATION:
; APPLICANT: Lockridge, Jeffery D.
; APPLICANT: Watkins, Oksana
; TITLE OF INVENTION: Butyrylcholinesterase Variants and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: P-IX 4143
; CURRENT APPLICATION NUMBER: US/09/748,739A
; CURRENT FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 574
; TYPE: PRT

Query Match      16.8%; Score 482; DB 14; Length 574;
Best Local Similarity 30.4%; Pred. No. 7.8e-37;
Matches 162; Conservative 78; Mismatches 193; Indels 100; Gaps 23;

Qy 22 GDTITGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGSLNGQKFTSYGSPSCMOONPEGT 81
Db 22 GGTVT-----AFLGIPYAPQPLRLRFRKPKQSLTKWSNINWATKYANSC-YQNTDQS 72
Qy 82 FEENLGTALDLVMSQVFOAVLPQSEDCLTINVRP-PGKAGANLPMVLWIFGGGPEI 140
Db 73 FPGFLG-----SEWNPNTSELSDCLYLVNWIAPKPK---NATVMIWYGGGFPQT 120
Qy 141 GSPITFPPAQMVTKSVLMGK-----HIHVAVNRYRVASWGLAGDDIKABSGSNAGLKD 194
Db 121 GTSSL-----FVYDGGKFLARVERVI VVSMNRYRGALGFLALSE-NPEAPGNMGLFD 170
Qy 195 QRLGMQWADNIAGFGDPKSVTIFGESAGSMVSLCHLIWNGDNTYKGKPLFRAGIMQS 254
Db 171 QQLALQWVQKNIAAFGNGPRSVTLFGESAGAAVSLSHLL-----SPRSQPLFTRAILQS 224
Qy 255 GAMVSPDPVGTYGNEIYDLFVSS-AGCG--SASDKLACLRSASSDTLLDATTNTPGFLA 311
Db 225 GSSNAPWAVTSLYEARNRTLTLAKRMGCSRDNEMETIKCLDKDPQERIL---LNEVFVVP 281
Qy 312 YSS-LRLSYLPRPDGKNITDDMYKLVRDGYKASVPVLIQNDRTGTF-----GLSSLN 364
Db 282 YDTLLSVNFGTVDGDFLTDPDTLLQLQGFQKRTQILVNVNKDEGTAFVYGAFGFSKDN 341
Qy 365 --VTTNAQARAYFKQSFHASDABIDTLMAAY-----PQDITQSPDPTGVNLALT 414
Db 342 NSIITRKEFGQGLKIPFPRVSEFGRESILFHYMDWLDQRAENVREALDDVVGVDYNIICP 401
Qy 415 --QPKRISAVLGDIAFHARRYFLNHFQGGTKYSFSLKQSLGSLPIMGTTHANDIHWQDY 472
Db 402 ALETRFKFSELGNDADF---YYPEH-----RSTKLPWPPEWGMVHGYEI---EFV 445
Qy 473 LG---SGSVIYNNA-----FIATATDLPN--TAGLLVWMPKYTSOSSQ 510
Db 446 FGLPLERRVNYTRABEILSRKIMKRWANFAKGNPNGTQNNSTWPFVKSTEQ 498

RESULT 12
US-10-032-233-48
; Sequence 48, Application US/10032233
; Publication No. US20030153062A1
; GENERAL INFORMATION:
; APPLICANT: Watkins, Jeffery D.
; APPLICANT: Hancock, James D.
; TITLE OF INVENTION: Butyrylcholinesterase Variants with
; TITLE OF INVENTION: Increased Catalytic Efficiency and Methods of Use
; FILE REFERENCE: P-IX 4642
; CURRENT APPLICATION NUMBER: US/10/032,233
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Equus caballus
US-10-032-233-48

Query Match      16.8%; Score 482; DB 14; Length 574;
Best Local Similarity 30.4%; Pred. No. 7.8e-37;
Matches 162; Conservative 78; Mismatches 193; Indels 100; Gaps 23;

Qy 22 GDTITGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGSLNGQKFTSYGSPSCMOONPEGT 81
Db 22 GGTVT-----AFLGIPYAPQPLRLRFRKPKQSLTKWSNINWATKYANSC-YQNTDQS 72
Qy 82 FEENLGTALDLVMSQVFOAVLPQSEDCLTINVRP-PGKAGANLPMVLWIFGGGPEI 140
Db 22 GGTVT-----AFLGIPYAPQPLRLRFRKPKQSLTKWSNINWATKYANSC-YQNTDQS 72
Qy 82 FEENLGTALDLVMSQVFOAVLPQSEDCLTINVRP-PGKAGANLPMVLWIFGGGPEI 140
```

```
Db 73 FPGFLG-----SEMMNPTELSEDCILYNVWIPAPKPK-----NATVMIWYGGGFOT 120
Qy 141 GSPTIFPPAQMTKSVLMGK-----HIIHVAVNYRVASWGLFAGDDIIKAGSGNAGLKD 194
Db 121 GTSSL-----PYDGGFLARVERVIVVSMYRVGALGFALSE-NPEAFGNMGLFD 170
Qy 195 QRLGMOWADNIAGFGDPSKVTIFGESAGSMVLCILWINDGNTYKKGKPLFRAGIMOS 254
Db 171 QQLALQWQKNIAAFQGNPRSVTLFGEAGASVSLHLL-----SPRSQPLETRAILQS 224
Qy 255 GAWPSPDVGDTYGNIEIYDLFVSS-AGCG--SASDKLACLRSSASDTLLDATNNTPGFLA 311
Db 225 GSSNAPWAVTSLYEARNRITLAKRMGCSRDNEMIKCLRDQDPOEIL---LNEVFVVP 281
Qy 312 YSS-LRLSYLPRDGNKNIITDDMKYLRDQKYASVPVLIQDQDEGTIF-----GLSSLN 364
Db 282 YDTLLSVNFGPTVDGFLTDMPTLLQLGQKRTQILVGVNKGDEGTAFVLYGAPGFSKDN 341
Qy 365 --VTTNAQARAYFKQSFIIHASDAEIDTMAAY-----PDITOGSPDPTGVLNALTP 414
Db 342 NSIITRKEFQEGKIFPPRVSEFGRSILFHYMDLDDQRAENYREALDDVGDYNIICP 401
Qy 415 --QFKRISAVIGDLAFIHARRYFLNHFQGTGYKYSFLSKOLSGLPIMGTTHANDIVMDYL 472
Db 402 ALEFTRKFSELGNDAFF---YFHEH-----RSTKLPPWPEWGMVHGVEI---EFV 445
Qy 473 LG---SGSVIYNNA-----FIATATDLDPN-TAGLLVNWPKYTSSSQ 510
Db 446 FGLPLERRVNYTRAEILSRIMKRWANFAKYNPNGTQNNSTRWPFVKSTEQ 498
```

## RESULT 13

```
US-10-324-466-48
; Sequence 48, Application US/10324466
; Publication No. US20040121970A1
; GENERAL INFORMATION:
; APPLICANT: Watkins, Jeffrey D.
; TITLE OF INVENTION: Butyrylcholinesterase Variant
; TITLE OF INVENTION: Polypeptides with Increased Catalytic Efficiency and Methods
; FILE REFERENCE: P-IX 5555
; CURRENT APPLICATION NUMBER: US/10/324,466
; PRIOR FILING DATE: 2002-12-20
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Equus caballus
US-10-324-466-48
```

```
Query Match 16.8%; Score 482; DB 16; Length 574;
Best Local Similarity 30.4%; Pred. No. 7.8e-37;
Matches 162; Conservative 78; Mismatches 193; Indels 100; Gaps 23;

Qy 22 GDTITGLNALINEAFIIPPAEPVGNLRFKDPVPYSGSLNGOKFTSYGFSQMQNPEGT 81
Db 22 GGTVT-----AFLGIPYAQPLGLRFPKPKQSLTKWSNIWNATKYANSC-YQNTDQS 72
Qy 82 FEENLGKALTDLVMQSKVFOAVLPQSEDCLTINVRP-PCTKAGANLPVLMWIFGGGFEI 140
Db 73 FPGFLG-----SEMMNPTELSEDCILYNVWIPAPKPK-----NATVMIWYGGGFOT 120
Qy 141 GSPTIFPPAQMTKSVLMGK-----HIIHVAVNYRVASWGLFAGDDIIKAGSGNAGLKD 194
Db 121 GTSSL-----PYDGGFLARVERVIVVSMYRVGALGFALSE-NPEAFGNMGLFD 170
Qy 195 QRLGMOWADNIAGFGDPSKVTIFGESAGSMVLCILWINDGNTYKKGKPLFRAGIMOS 254
Db 171 QQLALQWQKNIAAFQGNPRSVTLFGEAGASVSLHLL-----SPRSQPLETRAILQS 224
Qy 255 GAWPSPDVGDTYGNIEIYDLFVSS-AGCG--SASDKLACLRSSASDTLLDATNNTPGFLA 311
Db 225 GSSNAPWAVTSLYEARNRITLAKRMGCSRDNEMIKCLRDQDPOEIL---LNEVFVVP 281
Qy 312 YSS-LRLSYLPRDGNKNIITDDMKYLRDQKYASVPVLIQDQDEGTIF-----GLSSLN 364
Db 282 YDTLLSVNFGPTVDGFLTDMPTLLQLGQKRTQILVGVNKGDEGTAFVLYGAPGFSKDN 341
```

```
Db 171 QQLALQWQKNIAAFQGNPRSVTLFGEAGASVSLHLL-----SPRSQPLETRAILQS 224
Qy 255 GAWPSPDVGDTYGNIEIYDLFVSS-AGCG--SASDKLACLRSSASDTLLDATNNTPGFLA 311
Db 225 GSSNAPWAVTSLYEARNRITLAKRMGCSRDNEMIKCLRDQDPOEIL---LNEVFVVP 281
Qy 312 YSS-LRLSYLPRDGNKNIITDDMKYLRDQKYASVPVLIQDQDEGTIF-----GLSSLN 364
Db 282 YDTLLSVNFGPTVDGFLTDMPTLLQLGQKRTQILVGVNKGDEGTAFVLYGAPGFSKDN 341
Qy 365 --VTTNAQARAYFKQSFIIHASDAEIDTMAAY-----PDITOGSPDPTGVLNALTP 414
Db 342 NSIITRKEFQEGKIFPPRVSEFGRSILFHYMDLDDQRAENYREALDDVGDYNIICP 401
Qy 415 --QFKRISAVIGDLAFIHARRYFLNHFQGTGYKYSFLSKOLSGLPIMGTTHANDIVMDYL 472
Db 402 ALEFTRKFSELGNDAFF---YFHEH-----RSTKLPPWPEWGMVHGVEI---EFV 445
Qy 473 LG---SGSVIYNNA-----FIATATDLDPN-TAGLLVNWPKYTSSSQ 510
Db 446 FGLPLERRVNYTRAEILSRIMKRWANFAKYNPNGTQNNSTRWPFVKSTEQ 498
```

## RESULT 14

```
US-10-413-432-48
; Sequence 48, Application US/10413432
; Publication No. US20040120939A1
; GENERAL INFORMATION:
; APPLICANT: Watkins, Jeffrey D.
; TITLE OF INVENTION: Butyrylcholinesterase Variant
; TITLE OF INVENTION: Polypeptides with Increased Catalytic Efficiency and Methods
; FILE REFERENCE: P-IX 5510
; CURRENT APPLICATION NUMBER: US/10/413,432
; CURRENT FILING DATE: 2003-04-11
; PRIOR APPLICATION NUMBER: US 10/324,466
; PRIOR FILING DATE: 2002-12-20
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Equus caballus
US-10-413-432-48
```

```
Query Match 16.8%; Score 482; DB 16; Length 574;
Best Local Similarity 30.4%; Pred. No. 7.8e-37;
Matches 162; Conservative 78; Mismatches 193; Indels 100; Gaps 23;

Qy 22 GDTITGLNALINEAFIIPPAEPVGNLRFKDPVPYSGSLNGOKFTSYGFSQMQNPEGT 81
Db 22 GGTVT-----AFLGIPYAQPLGLRFPKPKQSLTKWSNIWNATKYANSC-YQNTDQS 72
Qy 82 FEENLGKALTDLVMQSKVFOAVLPQSEDCLTINVRP-PCTKAGANLPVLMWIFGGGFEI 140
Db 73 FPGFLG-----SEMMNPTELSEDCILYNVWIPAPKPK-----NATVMIWYGGGFOT 120
Qy 141 GSPTIFPPAQMTKSVLMGK-----HIIHVAVNYRVASWGLFAGDDIIKAGSGNAGLKD 194
Db 121 GTSSL-----PYDGGFLARVERVIVVSMYRVGALGFALSE-NPEAFGNMGLFD 170
Qy 195 QRLGMOWADNIAGFGDPSKVTIFGESAGSMVLCILWINDGNTYKKGKPLFRAGIMOS 254
Db 171 QQLALQWQKNIAAFQGNPRSVTLFGEAGASVSLHLL-----SPRSQPLETRAILQS 224
Qy 255 GAWPSPDVGDTYGNIEIYDLFVSS-AGCG--SASDKLACLRSSASDTLLDATNNTPGFLA 311
Db 225 GSSNAPWAVTSLYEARNRITLAKRMGCSRDNEMIKCLRDQDPOEIL---LNEVFVVP 281
Qy 312 YSS-LRLSYLPRDGNKNIITDDMKYLRDQKYASVPVLIQDQDEGTIF-----GLSSLN 364
Db 282 YDTLLSVNFGPTVDGFLTDMPTLLQLGQKRTQILVGVNKGDEGTAFVLYGAPGFSKDN 341
```

QY 365 --VTTNAQARAYFKQSIHASDAEIDTILMAAY-----PQDITQSPDPTGVNLATP 414  
 Db 342 NSIITRKEFOEGLKIFFERVSEFGRESILFHMYDMLDDQRAENTYREALDDVVGDYNIICP 401  
 QY 415 --QFKRISAVLGDLAFIHARRYFLNHFQGGTKYSFLSKQLSGLPIMGTFFHANDIWMODYL 472  
 Db 402 ALETRKFSGLNDAFF-----YYPEH-----RSTKLPPEWGMVHGVEI---EFV 445  
 QY 473 LG---SGSVYNNNA-----FIATFDLDPN-TAGLLVNWPKYTSSSQ 510  
 Db 446 FGLPLERRVNYTRAEIILRSIMKRWANFAKYGPNGTQNNSTRWPVPKSTEQ 498

RESULT 15

US-09-748-739A-23  
 ; Sequence 23, Application US/09748739A  
 ; Patent No. US20020119489A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lockridge, Oksana  
 ; APPLICANT: Watkins, Jeffrey D.  
 ; TITLE OF INVENTION: Butyrylcholinesterase Variants and  
 ; TITLE OF INVENTION: Methods of Use  
 ; FILE REFERENCE: P-IX 4143  
 ; CURRENT APPLICATION NUMBER: US/09/748,739A  
 ; CURRENT FILING DATE: 2000-12-06  
 ; NUMBER OF SEQ ID NOS: 31  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 23  
 ; LENGTH: 574  
 ; TYPE: PRT  
 ; ORGANISM: Rattus sp.  
 US-09-748-739A-23

Query Match 16.7%; Score 477; DB 9; Length 574;  
 Best Local Similarity 30.7%; Pred. No. 2.4e-36;  
 Matches 164; Conservative 77; Mismatches 190; Indels 104; Gaps 25;

QY 22 GDTITGLNAINEAFGLIPFAEPVGNLRPKDPVPGYSLNGOKFTSYGSCMOQNPEGT 81  
 Db 22 GGTVT-----AFLGIPYAPQPLGSLRFRKPKQPLNKWPDPVYNATKYANSC-YQNIQDA 72  
 QY 82 FEENLGKTALDLMQSKVFOAVLPQSEDCLTINVRP-PGTRKAGANFLVMLWIFGGFPEI 140  
 Db 73 FPGFG-----SEMNNTNLSEDCLYLNVWIPVPKP---NATVMVWYGGGFQT 120  
 QY 141 GSPTIFPPAQMTKSVLMGK-----HIIHVAVNYRVASMGFLAGDDIKAECSGNAGLKD 194  
 Db 121 GTSSL-----PVYDGKELTRVERVIVVSMYRVGALGFAPPG-NSEAPGNMGLFD 170  
 QY 195 QRLGOMVADNIAGGDPKSVTIIFGESAGSMVLCHLIWNGDNTYKGLPLFRAGIMQS 254  
 Db 171 QQLALQMTQRIAAFGGNPKSVTLFGESAGASVSLHLL-----CPOSYPLFTRAILES 224  
 QY 255 GAMPSPDVGDTYGNIEIYDLFVSSA---GCGSASDK--LACLRSSADTLDDATNTPGF 309  
 Db 225 GS--SNAPWAKHPEARNRTITLAKFIGCSKENEKEIITCLRSKDPQELL--LNEKLV 279  
 QY 310 LAYSSLR-LSVLPDPGNITDDMYKLVDRDKYASVPVVIIGDQNDGHIIF-----GLSS 362  
 Db 280 LPDSIRSINFGPTVDGDFLTMPHTLQLGKVKTAQLVGVNKDEGTAFIVYGAPGFSK 339  
 QY 363 LN--VTTNAQARAYFKQSIHASDAEIDTILMAAYPQDITQSP-----FD--TGVNLAL 412  
 Db 340 DNDSLITRREFQEGGLNMTFPGVSSLGKEAILFYVDNLQDQTPVYREAFDDIIGDYNII 399  
 QY 413 TP--QFKRISAVLGDLAFIHARRYFLNHFQGGTKYSFLSKQLSGLPIMGTFFHANDIWMQD 470  
 Db 400 CPALFTKKFAELEINAF--YYPEH-----RSTKLPPEWGMVHGVEI---E 443  
 QY 471 YLLG---SGSVYNNNAFIATFDL-----DPN-TAGLLVNWPKYTSSSQ 510  
 Db 444 FVFGFLPLERRVNYTRAEIILRSIMKRWANFAKYGPNGTQNNSTRWPVPVFTSTEQ 498

Search completed: July 29, 2004, 10:56:20  
 Job time : 52 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 10:48:09 ; Search time 19 Seconds  
(without alignments)  
1486.285 Million cell updates/sec

Title: US-09-943-857-4

Perfect score: 2864

Sequence: 1 SMNSRGPAGRLGVSPTAKLA.....DNFRTAGYDALMTNPSFFV 547

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

1: /cgn2\_6/prodata/2/iaa/5A COMB.pap.\*

2: /cgn2\_6/prodata/2/iaa/5B COMB.pap.\*

3: /cgn2\_6/prodata/2/iaa/6A COMB.pap.\*

4: /cgn2\_6/prodata/2/iaa/6B COMB.pap.\*

5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pap.\*

6: /cgn2\_6/prodata/2/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1107	38.7	563	3	US-08-362-525-12
2	672.5	23.5	574	4	US-10-023-515-4
3	491.5	17.2	575	1	US-08-348-920-1
4	488.5	17.1	575	1	US-08-348-920-2
5	454	15.9	614	3	US-08-446-100-25
6	452	15.8	600	2	US-08-370-156-4
7	452	15.8	600	3	US-08-814-095-4
8	452	15.8	600	3	US-08-975-084-1
9	452	15.8	602	6	5215909-11
10	452	15.8	614	1	US-07-732-962A-2
11	452	15.8	614	2	US-08-370-156-2
12	452	15.8	614	3	US-08-446-100-19
13	452	15.8	614	3	US-08-446-100-21
14	452	15.8	614	3	US-08-814-095-2
15	452	15.8	614	5	PCT-US92-06106-2
16	452	15.8	617	3	US-08-370-156-6
17	452	15.8	617	3	US-08-814-095-6
18	451	15.7	602	3	US-08-446-100-1
19	451	15.7	602	3	US-08-446-100-24
20	451	15.7	602	3	US-09-334-489-3
21	451	15.7	602	3	US-09-334-489-4
22	449.5	15.7	572	2	US-08-932-376A-2
23	449	15.7	602	3	US-08-446-100-13
24	448	15.6	602	3	US-08-446-100-11
25	447	15.6	602	3	US-08-446-100-3
26	447	15.6	602	3	US-08-446-100-4
27	447	15.6	602	3	US-08-446-100-5

28 447 15.6 602 3 US-08-446-100-12 Sequence 12, Appl  
29 447 15.6 602 3 US-08-446-100-15 Sequence 15, Appl  
30 447 15.6 602 3 US-08-446-100-16 Sequence 16, Appl  
31 447 15.6 602 3 US-08-446-100-17 Sequence 17, Appl  
32 447 15.6 602 3 US-08-446-100-18 Sequence 18, Appl  
33 447 15.6 614 3 US-08-446-100-22 Sequence 22, Appl  
34 447 15.6 614 3 US-08-446-100-23 Sequence 23, Appl  
35 446 15.6 602 3 US-08-446-100-8 Sequence 8, Appl  
36 445 15.5 602 3 US-08-446-100-14 Sequence 14, Appl  
37 444.5 15.5 544 2 US-08-932-376A-4 Sequence 4, Appl  
38 444.5 15.5 635 6 5215909-10 Patent No. 5215909  
39 444 15.5 614 3 US-08-446-100-20 Sequence 20, Appl  
40 443 15.5 573 6 5215909-12 Patent No. 5215909  
41 443 15.5 602 3 US-08-446-100-2 Sequence 2, Appl  
42 443 15.5 602 3 US-08-446-100-6 Sequence 6, Appl  
43 442 15.4 602 3 US-08-446-100-7 Sequence 7, Appl  
44 442 15.4 602 3 US-08-446-100-9 Sequence 9, Appl  
45 442 15.4 602 3 US-08-446-100-10 Sequence 10, Appl

#### ALIGNMENTS

##### RESULT 1

US-08-362-525-12

; Sequence 12, Application US/08362525

; Patent No. 6027910

; GENERAL INFORMATION:

; APPLICANT: KLIS, FRANCISCUS M.

; APPLICANT: SCHREUDER, MAARTEN P.

; APPLICANT: TOSCHKA, HOLSER Y.

; APPLICANT: VERRIPS, CORNELIS T.

; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE

; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION

; TITLE OF INVENTION: PROTEIN

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.

; STREET: 1100 New York Avenue, N.W.

; CITY: Washington

; STATE: D. C.

; COUNTRY: U.S.A.

; ZIP: 20005-3918

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/362,525

; FILING DATE: 04-JAN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 92202080.5

; FILING DATE: 08-JUL-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 92203899.7

; FILING DATE: 14-DEC-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/EP93/01763

; FILING DATE: 07-JUL-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: KOKULIS, PAUL N.

; REGISTRATION NUMBER: 16,773

; REFERENCE/DOCKET NUMBER: 213289/T7020 (V)

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 861-3000

; TELEFAX: (202) 822-0944

; TELEX: 6714627 CUSH

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 563 amino acids

; TYPE: amino acid

```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-362-525-12

Query Match      38.7%; Score 1107; DB 3; Length 563;
Best Local Similarity 43.4%; Pred. No. 5.4e-100;
Matches 240; Conservative 73; Mismatches 196; Indels 44; Gaps 9;

Qy 9  GRLGSVTAKLANGDTTGLNAINAEPLGTIPFAEPVGNLRFKDPYPSGSLNGQKETS 68
Db 16 GTLAQAATAVLNGNEVSGVLEKGVDTFKGPIFADPPVGDRLFKHPQPFCTSGYQLKAND 75
Qy 69 YGPMCOQNPBGTE-----ENLGKTAIDLVMQSKVFQAVLPQSEDCLTINV 115
Db 76 FSSACQLDPGNASLLDKVVLGKLIPDLNLRGLPDMA-----QGSVSNEDCLYINV 129
Qy 116 VRPPTKAGANLPVLMITFGGFEIGSTPIPPAQMTKSVLMGKHIIHVAENVYRVSAG 175
Db 130 FRPAGTRKPAKLPVMMIYGGAFVFGSSASYPGNGYVKESVEMGQPVVVFVSINRYRTGPG 189
Qy 176 FLAGDDIKAECSGNAGLKQRLGMQWVADNIAGGDPKSYTIIFGESAGMSVLCHLIWN 235
Db 190 FLGGDAITAEGNTNAGLHDQRKLEWMSDNIANFGDDPKYMIIFGESAGAMVAHQLVAY 249
Qy 236 DGDNTYKGPPLFRAGIMOSGAMVPDSDPVDGTGYGNEIYDLFVSSAGC---GSASDKLACL 292
Db 250 GGDNTYNGKQLFHSNAILQSGPLPYFDSTSVGPESAYSRAFYAGCDTSASDNDTLACL 309
Qy 293 SASDSTLLDATNN-----TGFPLAYSSLRSLYLPDPGKNITDDMYKLVRDGYAS 343
Db 310 SKSSDVLHSAQNSYDLKDLFLGLPQLFG-----PRPDGNIIPDAAYELYSRGYAK 362
Qy 344 VPVIIGQNDGRTTGLSSLVNTNAQARAYFKQSFHASDAEIDTLMAAYPQDITGSP 403
Db 363 VPYITGQNDGRTTGLAPVAIATTPPHVKKWLKICQASDASDLRDLVLSLPGSWSESP 422
Qy 404 PDTGVNALTQPKRISAVLGLAFIHARRYFLNHFQGTGKYFSLSKOLSLG-L-PIMGTFH 462
Db 423 FRTGLNALTQPKRISAVLGLAFIHARRYFLNHFQGTGKYFSLSKOLSLG-L-PIMGTFH 482
Qy 463 ANDIVQDYL-LGSGSVIYNNAFATADLPNTAGLLVNNPKYTSQSGSNLMMINAL 521
Db 483 GSDLLFQYVVDLGPSSA-YRRYFISFANHDPNVGNTLQQDMYT---DAGKEMLIQHMI 538
Qy 522 GLYTCKDNFRTAG 534
Db 539 GNSMRTDDFRLEG 551

RESULT 2
US-10-023-515-4
; Sequence 4, Application US/10023515
; Patent No. 6664091
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Silos-Santiago, Inmaculada
; TITLE OF INVENTION: 53010. A NOVEL HUMAN CARBOXYLESTERASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREOF
; FILE REFERENCE: 10448-122001
; CURRENT APPLICATION NUMBER: US/10/023,515
; CURRENT FILING DATE: 2001-12-18
; PRIOR FILING DATE: 2000-12-18
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/279,508
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence

Query Match      23.5%; Score 672.5; DB 4; Length 574;
Best Local Similarity 34.9%; Pred. No. 3.3e-57;
Matches 195; Conservative 76; Mismatches 195; Indels 93; Gaps 18;

Qy 35 AFLGIPFAEPVGNLRFKDPVY-----SGSLNGQKFTSYGSCMQQNPGEFTFENLGT 90
Db 19 SFLGIPFAEPVGNLRFKAPQYKPPWSDVLDA---TKYPSCQLQDDDFGFSLSDL-KVA 74
Qy 91 LDLVMQSKVQAVLPQSEDCLTINVPPGPKAGANLPVLMIFGGGFEIGTSPIFPQAQ 150
Db 75 LKMLSLGWNKLVGLKJSEDCLYLVYTPKNTKPSKLPMVMVHHGGGFMFGSGHSLPUL 134
Qy 151 MVTKSLVLMGKHIIHVAENVYRVSAGVFLAGDDIKAECSGNAGLKQRLGMQWVADNIAGFG 210
Db 135 YDGESLAREGNIVVSYNYRLGFLGFLSTGDDKLPKSGNYGLLDLQRLALKWVQDNIAFG 194
Qy 211 GDPKSVTIIFGESAGMSVLCHLIWNGDNTYKGPLFRAGIMQSGA-----MVPSPDPVDG 265
Db 195 GDPNSVTIFGESAGAAVSLLLLISNGGDNPPSPKGLFHRAISQSGALSPWAIQSESNAR 254
Qy 266 TYGNEIYDLFVSSAGC---SASDKLACLRSSASDNLADATNTTPEFLAYS-----SLRLS 318
Db 255 GRAKELARLL---GCNETSSSELLDCLRSKSAELLEATRS---FLLFEVVPPLFLFLA 307
Qy 319 YLPRPDGKN-----ITDDMYKLVRDGYASVPVIIGQNDGRTTGLSSLVNTNAQAR- 372
Db 308 FGPVVDGDDAPEAFIPEDPELIKEGKFAVPLYIGTKDEGGYFAAMLNASSKGEDEL 367
Qy 373 -----AYFKQSFHASDA-----EITLMAAYPQDITGSPFDTGVNLNATP 414
Db 368 KKETNPDMVLELLKYLFFYASEALNKDMDDLDKVLKYPGDVDDFS-----YES 418
Qy 415 QPKISAVLGLAF-----IHARR-----YFLNHFQGTGKYFSLSKOLSLGSLPIM 458
Db 419 RKNLQDMLTDLLEPKCPTRVAAADLHAKHGSPVYAVFDHPASFGIGQFLAKRVDP-BFG 477
Qy 459 GTFHANDI--VWQDYL-----GSGSVIYNNAFATADLPNTAGLLVNNPKYTSQSGSNLMMINAL 502
Db 478 GAVGDEIRFFVFGNPLLEKQYKATEBEEKSSSKTMMYWANFAKTGNPNNGTSNGLVVM 537
Qy 503 PKYTSSSQSGSNLMMINAL 521
Db 538 PKYTSEQKYSLLILATTI 556

RESULT 3
US-08-348-920-1
; Sequence 1, Application US/08348920
; Patent No. 5695750
; GENERAL INFORMATION:
; APPLICANT: Doctor, Bhupendra P.
; APPLICANT: Maxwell, Donald
; APPLICANT: Saxena, Ashima
; APPLICANT: Radic, Zoran
; APPLICANT: Taylor, Palmer
; TITLE OF INVENTION: Compositions for Use to Deactivate
; TITLE OF INVENTION: Organophosphates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John F. Moran
; STREET: Off. of Command Judge Adv., HQ USAMRDC, Fort
; STREET: Detrick
; CITY: Frederick
; STATE: MD
; COUNTRY: US
; ZIP: 21702-5012
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```



```

/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/348,920
/ FILING DATE: 25-NOV-1994
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Hendricks, Glenma
/ REGISTRATION NUMBER: 32,535
/ REFERENCE/DOCKET NUMBER: doc348,920
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 619-7807
/ TELEFAX: 301-619-7714
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 575 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: unknown
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: internal
/
US-08-348-920-1

Query Match 17.2%; Score 491.5; DB 1; Length 575;
Best Local Similarity 30.4%; Pred. No. 2.2e-39;
Matches 164; Conservative 75; Mismatches 195; Indels 105; Gaps 24;

QY 22 GDTITGLNALINEAFIIPFAEPVGNLRFKDP---VPYSGSLNGKFTSYGSCMQQNP 78
Db 17 GTRVPVLSHSIS-AFUGIPFAEPVGNMFRPRPEPKKPMWGVWNASTY-----PNNCQQYV 71
QY 79 EGTFEENLGTALDLVMQSKVFQAVLPQSEDCLTINV---VRPPGTKAGANLPVMLWIF 134
Db 72 DEQFGFGSG-----SEMNPNREMSDCLYLNIVWPSRPKST-----TVMVWIY 116
QY 135 GGGFEIGSTIIPPAOMVTKSVLMGXH-----IIHVAVNYRVASWGFLA--GDDIKAE 186
Db 117 GGGFYSGSSTL-----DVYNGKYLAYTEBVLVLSYRVGAFGLALHGSQ---EA 164
QY 187 SGNAGLKDORLGOMQWADNIAGGDPKSVTIIFGESAGSMLVCHLIWNGDNTYKGP 246
Db 165 PGNVGLDQRMALQVHDNIQFFGDPKVTIIFGESAGGASVGMHIL-----SPGSRDL 218
QY 247 FRAGIMQSGAMVPSDP---VDGTGNGEYDLFVSSAGCGSASDK--LACLRASSTL 301
Db 219 FRAILQSGS--PNCPSASVSVAEGRRAVELGRNLNCLNLSDEELIHLREKKPQELID 276
QY 302 ATNNTPGFLAYSSL-RLSVLPDPDKNITDDMYKLVDRGKYASVPVVIIGQNDDEGTF 360
Db 277 VEWNV---LPFDSIFRFSFVPVIDGEFFFTSLESMLNSGNFKKTKQLLVGNKDEGSF 333
QY 361 -----SSLNVTNAQARAYFKQSFHASDAEIDTLMAAYPQDITQSGPDTGVLN 412
Db 334 YGAPGSKDSEKISREDPMWGVKLSVPHANDLGLDAVTLQY-----TDMWDDNNGIK 388
QY 413 TPQFKRISAVLGLDAFIHARRYFLN---HFQGGTKYSFLSKQLSGL---PIMGT 466
Db 389 ----DGLDDIVGDHNVICPLMHFVNKYTKFGNGTYLYFENHRASNLVPEWGVHGYEI 444
QY 467 VMQDYLLGGSGSVIYNN-----AFTAFATDLDNPTA-GLLVNNPKYTSQS 510
Db 445 ---EFVFGFLPLKELNYTAEAEALSRRIMHYWATFAGTGNPNPHSQSKWPLFTTKEQ 500

RESULT 4
US-08-348-920-2
/ Sequence 2, Application US/08348920
/ Patent No. 5695750
/ GENERAL INFORMATION:
/ APPLICANT: Doctor, Bhupendra P.
/ APPLICANT: Maxwell, Donald
/ APPLICANT: Saxena, Ashima
/ APPLICANT: Radic, Zoran

```

```

/ APPLICANT: Taylor, Palmer
/ TITLE OF INVENTION: Compositions for Use to Deactivate
/ NUMBER OF SEQUENCES: 2
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: John F. Moran
/ STREET: Off. of Command Judge Adv., HQ USAMRDC, Fort
/ CITY: Frederick
/ STATE: MD
/ COUNTRY: US
/ ZIP: 21702-5012
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ FILING DATE: 25-NOV-1994
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Hendricks, Glenma
/ REGISTRATION NUMBER: 32,535
/ REFERENCE/DOCKET NUMBER: doc348,920
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 619-7807
/ TELEFAX: 301-619-7714
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 575 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: unknown
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: internal
/
US-08-348-920-2

Query Match 17.1%; Score 488.5; DB 1; Length 575;
Best Local Similarity 30.2%; Pred. No. 4.4e-39;
Matches 163; Conservative 76; Mismatches 195; Indels 105; Gaps 24;

QY 22 GDTITGLNALINEAFIIPFAEPVGNLRFKDP---VPYSGSLNGKFTSYGSCMQQNP 78
Db 17 GTRVPVLSHSIS-AFUGIPFAEPVGNMFRPRPEPKKPMWGVWNASTY-----PNNCQQYV 71
QY 79 EGTFEENLGTALDLVMQSKVFQAVLPQSEDCLTINV---VRPPGTKAGANLPVMLWIF 134
Db 72 DEQFGFGSG-----SEMNPNREMSDCLYLNIVWPSRPKST-----TVMVWIY 116
QY 135 GGGFEIGSTIIPPAOMVTKSVLMGXH-----IIHVAVNYRVASWGFLA--GDDIKAE 186
Db 117 GGGFYSGSSTL-----DVYNGKYLAYTEBVLVLSYRVGAFGLALHGSQ---EA 164
QY 187 SGNAGLKDORLGOMQWADNIAGGDPKSVTIIFGESAGSMLVCHLIWNGDNTYKGP 246
Db 165 PGNVGLDQRMALQVHDNIQFFGDPKVTIIFGESAGGASVGMHIL-----SPGSRDL 218
QY 247 FRAGIMQSGAMVPSDP---VDGTGNGEYDLFVSSAGCGSASDK--LACLRASSTL 301
Db 219 FRAILQSGS--PNCPSASVSVAEGRRAVELGRNLNCLNLSDEELIHLREKKPQELID 276
QY 302 ATNNTPGFLAYSSL-RLSVLPDPDKNITDDMYKLVDRGKYASVPVVIIGQNDDEGTF 360
Db 277 VEWNV---LPFDSIFRFSFVPVIDGEFFFTSLESMLNSGNFKKTKQLLVGNKDEGSF 333
QY 361 -----SSLNVTNAQARAYFKQSFHASDAEIDTLMAAYPQDITQSGPDTGVLN 412
Db 334 YGAPGSKDSEKISREDPMWGVKLSVPHANDLGLDAVTLQY-----TDMWDDNNGIK 388
QY 413 TPQFKRISAVLGLDAFIHARRYFLN---HFQGGTKYSFLSKQLSGL---PIMGT 466

```





```
;
;
; REGISTRATION NUMBER: 38,972
; REFERENCE/DOCKET NUMBER: 2391.00082
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (248) 539-5050
; TELEFAX: (248) 539-5055
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-975-084-1

Query Match 15.8%; Score 452; DB 3; Length 600;
Best Local Similarity 29.1%; Pred. No. 1.8e-35;
Matches 156; Conservative 83; Mismatches 179; Indels 118; Gaps 26;

Qy 35 AFLGIPPAEPVGNLRPKDP--VPYSGSLGKQFTSYGSCMO-----QNPEGTFEEN 85
Db 62 AFLGIPPAEPVGNLRPKDP--VPYSGSLGKQFTSYGSCMO-----CYQVVDLYPGFEGTEMWN 118
Qy 86 LGKTALDLMQSKVFOAVLPQSEDCLTINVRPPTKAGANLPVLMIFPGGFGIGSPTI 145
Db 119 PNREL-----SEDCLYLN-VMTVPYRPTSPTPLVLMWYGGFGYSGASSL 161
Qy 146 PPPAQMTKSVLMGKHIIH-----VAVNRYVASWGLAGDDDIKABGSGNAGLQDRLGM 199
Db 162 -----DYYDGRFLVQAERTVLVSMNRYVGAFGLALPGSR-EAPGNVGLLDORLAL 211
Qy 200 QWADNLAGFGGDPKSVTIIFGESAGSMVLCILWINDGNTYKGLPLFRAGIMOSGAMVP 259
Db 212 QWQENVAAGGDPSTVTLFESAGAAVSVMHLL-----SPPERGLFHRAVLQSGA--P 263
Qy 260 SDPVDGTYG-----NEIYDLFVSSAGC-----GSADKLACLRASASDITLLDATNN 305
Db 264 NGPW-ATVGMGEARRATQLAHL-----VGCPPGGTGGNDTELVACLRTPAQVLV---NH 315
Qy 306 TPGFLAYSSL-RLSYLPRPDGKNTDDMYKLVROGKVASVPVIGDQNDGTFIP----- 358
Db 316 EHWVLPQESVFRFSFVPVVDGFLSDTPEALINAGDFHGLQVLGVGVKDEGSYFLVYGAP 375
Qy 359 GLSSLN--VTTNAQARAYFKQSFIHASDAEIDTLMAAY-----POD-----ITQGGPFDTVG 408
Db 376 GFSKDNESLISRAEFLAGVGVGVQVSDLAABAVVLHYTDMLHPEDPARLREALSDVVGD 435
Qy 409 LNALTPQPKRISAVGLDLAFIAR--RYFLNHFGQ-----GTKYSFLSKQLSGLP- 457
Db 436 HNVVCF-----VAQLAGRLAAQGARVYAVVFEHRASTLSWPLMGMVPHGYEIEFIFGIPLD 491
Qy 458 -MGTFHNDIWDQYLLGSGSVIYNNAFIATDLDNPA--GLLVNWPKYTSSSQ 510
Db 492 PSRNYTABEKIFAQRLM-----RYWANFARTGDPNEPRDPKAPQWPPYTAGAQ 539

RESULT 9
5215909-11
; Patent No. 5215909
; APPLICANT: SOREX, HERMONA
; TITLE OF INVENTION: HUMAN CHOLINESTERASE GENES
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/572,911
; FILING DATE: 15-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 87,724
; FILING DATE: 21-AUG-1987
; APPLICATION NUMBER: 875,737
; FILING DATE: 18-JUN-1986
; SEQ ID NO: 11:
; LENGTH: 602

5215909-11
Query Match 15.8%; Score 452; DB 6; Length 602;
Best Local Similarity 29.5%; Pred. No. 1.9e-35;
Matches 158; Conservative 82; Mismatches 150; Indels 106; Gaps 25;

Qy 22 GDTITGLNATINEAFILGIPFABPVGNLRPKDPVPYSGSLGKQFTSYGSCMOQNEPQT 81
Db 50 GGTVT-----AFLGIPVQAQPLGRLEKFKPQSLTKWSDIMWNAKYANSCC-QNIDQS 100
Qy 82 FEENLGTALDLMQSKVFOAVLPQSEDCLTINVRP--PGTKAGANLPVLMWIFGGGFEI 140
Db 101 FPGFHG-----SEMNPNPTDLSDCLYLNVWIPAPKEK---NATVLIWITGGGFGT 148
Qy 141 GSPTIFPPAQMTKSVLMGK-----HIIHVAVNRYVASWGLAGDDDIKABGSGNAGLKD 194
Db 149 GTSSL-----HVYDGKFLARVERVIVVSMNRYVGALGFLAWPG-NPEAPGNMGLFD 198
Qy 195 QRLGMQWADNLAGFGGDPKSVTIIFGESAGSMVLCILWINDGNTYKGLPLFRAGIMOS 254
Db 199 QQLALQWQKNIAFAFGNPKSVTLFESAGAAVSLSHL--SPGSHS-----LFTAILQS 252
Qy 255 GAMVPSDPVDGTY--GNEIYDLFVSSAGC--SASDKLACLRASASDITLLDATNTPGFL 310
Db 253 GSFNAPWAVTSLYEARNRTNL-AKLTGCSRENTEIILKLRNKDPOEIL---LNEAFV 308
Qy 311 AYSS-LRLSYLPRPDGKNTDDMYKLVROGKVASVPVIGDQNDGTFIP-----GLSSL 363
Db 309 PYGTPLSVNFGPTVDGDFLIDMPDILLELGGQFKTKQILVGVNKDEGTAFLVYGAPGSKD 368
Qy 364 N--VTTNAQARAYFKQSFIHASDAEIDTLMAAY-----PODITQGGPFDTVGLNALT 413
Db 369 NNSIITRKEFOGLKIFPPGVSEFGKESILPHYTDWVDQRPENYREALGDVWGDYDNFIC 428
Qy 414 P-QFKRISAVGLDLAFIARRYFLNHFGQTKYSFLSKQLSGLPIMCTPHANDIWMODY 471
Db 429 PALBETKFKSEWGNAPF---YIFEH-----RSSKLPPWFEGWVMGGEI---EF 472
Qy 472 LLG-----SGSVIYNNAFIATDLDNPA--TAGLLVWNPKYTSSSQ 510
Db 473 VFGLPLERRDNYTKAEILSRIV--KRWANFAKYGPNETQNNSTWPFVFKSTEQ 526

RESULT 10
US-07-732-962A-2
; Sequence 2, Application US/07732962A
; Patent No. 5248604
; GENERAL INFORMATION:
; APPLICANT: Fischer, Meir
; TITLE OF INVENTION: EXPRESSION OF ENZYMATICALLY ACTIVE
; TITLE OF INVENTION: RECOMBINANT HUMAN ACETYLCHOLINESTERASE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White, Esq.
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/732,962A
; FILING DATE: 19910722
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 39304/JPW/LSW
; TELECOMMUNICATION INFORMATION:
```

```
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 614 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-732-962A-2

Query Match      15.8%; Score 452; DB 1; Length 614;
Best Local Similarity 29.1%; Pred. No. 1.9e-35;
Matches 156; Conservative 83; Mismatches 179; Indels 118; Gaps 26;

QY 35 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGKFTSYGSPCMQ-----QNPEGTFEEN 85
Db 62 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGKFTSYGSPCMQ-----QNPEGTFEEN 118
QY 86 LGKTDLDLVMQSKVQAVLPQSEDCLTINNVPRPGTKAGANLPVLMWIFGGGFEIGSPTI 145
Db 119 PNREL-----SEDCLYLN-VWTPYPRPTSPTPLVWIIYGGGFYSGASSL 161
QY 146 FPPAQMTKSVLMGXHIIH-----VAVNYRVASWGLAGDDIKAEKSGNAGLKDORLGM 199
Db 162 -----DVIDGRFLVQAERTVLVSMNYRVGAFGLALPGSR-EAPGNVGLLDORLAL 211
QY 200 QWVADNIAFGDPSKVTIFGESAGSMSVLCILWINDGNTYKGLPFRAGIMQSGMVP 259
Db 212 QWQVENVAFAFGDPSVTILFESAGASVGMHLL-----SPPSRGLFHRAVLQSGA--P 263
QY 260 SDPVDTGYG-----NEIYDLFVSSAGC-----GSASDKLACLRSSASDTLLDATNN 305
Db 264 NGPW-ATVGMGARRRATOLAH-----VCGPPGGTGGNDTELVAQLTRPAQVLV---NH 315
QY 306 TPGFLAYSSL-RLSYLPRPDGKNIITDDMYKLVRDGYASVPVLIIGDQDEGTIF----- 358
Db 316 EWHVLPOESVFRFSFVVVDGDFLSDTPREALINAGDFHGLQVLGVVKGDEGSFVLVYGA 375
QY 359 GLSSLN--VTTNAQARAYFKQSFIIHASDAEIDTLMAAY-----PQD---ITQSPDPTGV 408
Db 376 GFSKONESLISRAEFLAGVRGVGPVQVSDLAEEAVLHYTDWLHPEDPARLREALSDVVDG 435
QY 409 LNALTPQFKRISAVLGDIAFIHAR--RYFLNHFGQ-----GTKYSFLSKOLSLGLPI- 457
Db 436 HNVVCP-----VAQLAGRLAAQGRVYVYVFEHRASTLSWPLWGVPHGVEIEFIFIGPLD 491
QY 458 -MGTFHANDIVWQDYLLGSGSVIYNNAFIAFATDLDNPTA--GLLVNWPKYTSSSQ 510
Db 492 PSRNYTAEEKIFAQRLM-----RYWANFARTGDPNEPRDPKAPQWPPYTAGAQ 539

RESULT 11
US-08-370-156-2
; Sequence 2, Application US/08370156
; Patent No. 5932780
; GENERAL INFORMATION:
; APPLICANT: Soreq, Hermona
; APPLICANT: Zakut, Haim
; APPLICANT: Shani, Moshe
; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
; TITLE OF INVENTION: ANTICHOLINESTERASE SUBSTANCES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reising, Rethington, Barnard & Perry
; STREET: P.O. Box 4390
; CITY: Troy
; STATE: Michigan
; COUNTRY: US
; ZIP: 48099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,156
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: P-307 (Mulford)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 689-3500
; TELEFAX: (810) 689-4071
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 614 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-370-156-2

Query Match      15.8%; Score 452; DB 2; Length 614;
Best Local Similarity 29.1%; Pred. No. 1.9e-35;
Matches 156; Conservative 83; Mismatches 179; Indels 118; Gaps 26;

QY 35 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGKFTSYGSPCMQ-----QNPEGTFEEN 85
Db 62 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGKFTSYGSPCMQ-----QNPEGTFEEN 118
QY 86 LGKTDLDLVMQSKVQAVLPQSEDCLTINNVPRPGTKAGANLPVLMWIFGGGFEIGSPTI 145
Db 119 PNREL-----SEDCLYLN-VWTPYPRPTSPTPLVWIIYGGGFYSGASSL 161
QY 146 FPPAQMTKSVLMGXHIIH-----VAVNYRVASWGLAGDDIKAEKSGNAGLKDORLGM 199
Db 162 -----DVIDGRFLVQAERTVLVSMNYRVGAFGLALPGSR-EAPGNVGLLDORLAL 211
QY 200 QWVADNIAFGDPSKVTIFGESAGSMSVLCILWINDGNTYKGLPFRAGIMQSGMVP 259
Db 212 QWQVENVAFAFGDPSVTILFESAGASVGMHLL-----SPPSRGLFHRAVLQSGA--P 263
QY 260 SDPVDTGYG-----NEIYDLFVSSAGC-----GSASDKLACLRSSASDTLLDATNN 305
Db 264 NGPW-ATVGMGARRRATOLAH-----VCGPPGGTGGNDTELVAQLTRPAQVLV---NH 315
QY 306 TPGFLAYSSL-RLSYLPRPDGKNIITDDMYKLVRDGYASVPVLIIGDQDEGTIF----- 358
Db 316 EWHVLPOESVFRFSFVVVDGDFLSDTPREALINAGDFHGLQVLGVVKGDEGSFVLVYGA 375
QY 359 GLSSLN--VTTNAQARAYFKQSFIIHASDAEIDTLMAAY-----PQD---ITQSPDPTGV 408
Db 376 GFSKONESLISRAEFLAGVRGVGPVQVSDLAEEAVLHYTDWLHPEDPARLREALSDVVDG 435
QY 409 LNALTPQFKRISAVLGDIAFIHAR--RYFLNHFGQ-----GTKYSFLSKOLSLGLPI- 457
Db 436 HNVVCP-----VAQLAGRLAAQGRVYVYVFEHRASTLSWPLWGVPHGVEIEFIFIGPLD 491
QY 458 -MGTFHANDIVWQDYLLGSGSVIYNNAFIAFATDLDNPTA--GLLVNWPKYTSSSQ 510
Db 492 PSRNYTAEEKIFAQRLM-----RYWANFARTGDPNEPRDPKAPQWPPYTAGAQ 539

RESULT 12
US-08-446-100-19
; Sequence 19, Application US/08446100
; Patent No. 6001625
; GENERAL INFORMATION:
; APPLICANT: Broomfield, Clarence A
; APPLICANT: Millard, Charles B
; APPLICANT: Lockridge, Oksana
; TITLE OF INVENTION: Site-Directed Mutagenesis of Esterases
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
```

ADDRESSEE: Hendricks and Assoc.  
STREET: 9669 A Main Street, P.O. Box 2509  
CITY: Fairfax  
STATE: VA  
COUNTRY: US  
ZIP: 22031  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446.100  
FILING DATE: 19-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hendricks, Glenna  
REGISTRATION NUMBER: 32,535  
REFERENCE/DOCKET NUMBER: broomfield  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 425-4250  
TELEFAX: (703) 425-2767  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 614 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: YES  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE: human esterases  
US-08-446-100-19

Query Match 15.8%; Score 452; DB 3; Length 614;  
Best Local Similarity 29.1%; Pred. No. 1.9e-35;  
Matches 156; Conservative 83; Mismatches 179; Indels 118; Gaps 26;

QY 35 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGKQFTSYGSPSCMQ-----QNPGTFTFEN 85  
DB 62 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGKQFTSYGSPSCMQ-----QNPGTFTFEN 118  
QY 86 LGKLTALDLVMSKVFQAVLPQSEDCLTINVRPPTKAGANLPVWLMTFGGFGFSGPTI 145  
DB 119 PNREL-----SEDCLYLN-VMTYPRPTSPTPVLVWYGGFGFSGASSL 161  
QY 146 FPPAQMTKSVLMGKHIIH-----VAVNYRVASWGLAGDDIKAEGSGNAGLKDQRLGM 199  
DB 162 -----DVYDGRFLVQAERTVLVSMNRYVGAFFLALPGSR-EAPGNVGLLDQRLAL 211  
QY 200 QWADNIAGFGDPSKVTIFGESAGSMVLCILWINDGNTYKGPPLFRAGIMOSGAMVP 259  
DB 212 QWQENVAAGFGDPTSVTLFGEAGAAASVGMHLL-----SPPSRGLFHRAVLQSGA--P 263  
QY 260 SDPVDGTG-----NEIYDLFVSSAGC-----QSASDKLACLRASASDILLDATNN 305  
DB 264 NGPW-ATVGMGEARRATQLAHL-----VGCPPGGTGGNDTELVAELRTPAQVLV--NH 315  
QY 306 TPGFLAYSSL-RLSYLPRPDGKNTDDMKYLRDGYKASVPVIGDQNDGTIF-----358  
DB 316 EWHVLPQESVPRFSPVVDGDFLSDTPEALINAGDFHGLQVLGVVVDKGSYFLVYAP 375  
QY 359 GLSSIN-VTTNAQARVFKOSFIHASDAEIDTLMAAY-----POD-----ITQSPFDGV 408  
DB 376 GFSKDNESLIRABFLAGRVGVVQVSDLAEEAVVHLHYDMLHPEDPARLREALSDVVGD 435  
QY 409 LNALTPQPKRISAVGLDLAFIAR--RYFLNHFG-----GTYKFSKQSLGSLPI- 457  
DB 436 HNVVCP-----VAQLAGLAAQAGARVYAVFEHRASTLSMPLWMGVPHGYEIEFFIGILD 491  
QY 458 -MGTHANDIVQDYLLGSGSVIYNNAFIATDLDPNTA--GLLVNWPKYTSSSQ 510

Db 492 PSRNYTAEKIFAQRLM-----RYWANPARTGDPNEPRDKAPQWPPYTAGAQ 539  
RESULT 13  
US-08-446-100-21  
Sequence 21, Application US/08446100  
Patent No. 6001625  
GENERAL INFORMATION:  
APPLICANT: Broomfield, Clarence A  
APPLICANT: Millard, Charles B  
APPLICANT: Lockridge, Oksana  
TITLE OF INVENTION: Site-Directed Mutagenesis of Esterases  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hendricks and Assoc.  
STREET: 9669 A Main Street, P.O. Box 2509  
CITY: Fairfax  
STATE: VA  
COUNTRY: US  
ZIP: 22031  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446.100  
FILING DATE: 19-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hendricks, Glenna  
REGISTRATION NUMBER: 32,535  
REFERENCE/DOCKET NUMBER: broomfield  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 425-4250  
TELEFAX: (703) 425-2767  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 614 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: YES  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE: human esterases  
US-08-446-100-21  
Query Match 15.8%; Score 452; DB 3; Length 614;  
Best Local Similarity 29.1%; Pred. No. 1.9e-35;  
Matches 156; Conservative 83; Mismatches 179; Indels 118; Gaps 26;  
QY 35 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGKQFTSYGSPSCMQ-----QNPGTFTFEN 85  
DB 62 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGKQFTSYGSPSCMQ-----QNPGTFTFEN 118  
QY 86 LGKLTALDLVMSKVFQAVLPQSEDCLTINVRPPTKAGANLPVWLMTFGGFGFSGPTI 145  
DB 119 PNREL-----SEDCLYLN-VMTYPRPTSPTPVLVWYGGFGFSGASSL 161  
QY 146 FPPAQMTKSVLMGKHIIH-----VAVNYRVASWGLAGDDIKAEGSGNAGLKDQRLGM 199  
DB 162 -----DVYDGRFLVQAERTVLVSMNRYVGAFFLALPGSR-EAPGNVGLLDQRLAL 211  
QY 200 QWADNIAGFGDPSKVTIFGESAGSMVLCILWINDGNTYKGPPLFRAGIMOSGAMVP 259  
DB 212 QWQENVAAGFGDPTSVTLFGEAGAAASVGMHLL-----SPPSRGLFHRAVLQSGA--P 263  
QY 260 SDPVDGTG-----NEIYDLFVSSAGC-----QSASDKLACLRASASDILLDATNN 305

Db 264 NGPW-ATVGMGEARRRATQLAHL-----VGPFGGTGGNDTELVAQLTRTPAQVLV---NH 315  
QY 306 TPGFLAYSSL-RLSYLPRPDGKNIITDDMYKLVRDGYASVVPVLIQDDEGTIF----- 358  
Db 316 EHWVLVQESVHRFSFVVPVVDGFLSTPEALINAGDFHGLQVLGVVVKDEGSFLVYGAP 375  
QY 359 GLSSLN--VTTNAQARAYFKQSFIIHASDAEIDTLMAAY-----PQD---ITQSPSPDTGV 408  
Db 376 GFSDKNESSLISRAEFLAGVRGVGPQVSDLAEEAVVLHYTDLHPEDPARLREALSDVVG 435  
QY 409 LNALTPQFKRISAVLGDIAFIHAR--RYFLNHPQG-----GTKYSLFKSLQSLGPI- 457  
Db 436 HNVVCP-----VAQLAGRLAAQGARVYAYVFEHRASTLSWPLMWGVPHGYEIEFIFGIP 491  
QY 458 -MGTTHANDIVWQDVLGSGSVIYNNAFIATATDLPNTA--GLLVNWPKYTSSSQ 510  
Db 492 PSRNTAEKIFAQRLM-----RYWANFARTGDPNEPRDKAPQWPPYTAGAQ 539  
  
RESULT 14  
US-08-814-095-2  
; Sequence 2, Application US/08814095  
; Patent No. 6025183  
; GENERAL INFORMATION:  
; APPLICANT: Soreq, Hermona  
; APPLICANT: Zakut, Haim  
; APPLICANT: Shani, Moshe  
; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR  
; TITLE OF INVENTION: ANTI-CHOLINESTERASE SUBSTANCES  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KOHN & ASSOCIATES  
; STREET: 30500 No. 6025183thwestern Highway, Suite 410  
; CITY: Farmington Hills  
; STATE: Michigan  
; COUNTRY: U.S.  
; ZIP: 48334  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/814,095  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Montgomery, Ilene N.  
; REGISTRATION NUMBER: 38,972  
; REFERENCE/DOCKET NUMBER: 2391.00066  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (248) 539-5050  
; TELEFAX: (248) 539-5055  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 614 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
US-08-814-095-2  
  
Query Match 15.8%; Score 452; DB 3; Length 614;  
Best Local Similarity 29.1%; Pred. No. 1.9e-35;  
Matches 156; Conservative 83; Mismatches 179; Indels 118; Gaps 26;  
  
QY 35 AFLGIFPAPPVGNLRFKDP---VPYSGSLNGQKETSYPSCMQ-----QNPEGTFEEN 85  
Db 62 AFLGIFAPPMPGRRFLPEPKQPSGVVDATTQSV---CQYVDTLYPGFEGEMWN 118  
QY 86 LGKLTALDLMQSKVFOAVLQSEDDCLTINVVRPGTKAGANLPMVLWIFGGGFEIGSPTI 145

Db 119 PNREL-----SEDCLYLN-VWTPYRPTSPTEVLVWLYGGGFYSGASSL 161  
QY 146 FPPAQMTKSLVMKGHIH-----VANNYRVASGFLAGDDIKAGBSGNAGLKDORLGM 199  
Db 162 -----DVTYDGRFLVQAERTVLVSMYRVRGAFGFLALPGSR-EAPGNVGLLDORLAL 211  
QY 200 QWVADNIAFGGDPGKVTIFGESAGSMSVLCHLIWNDGNTYKGPFLFRAGIMQSGAMVP 259  
Db 212 QWQENVAFAFGDPTSVTLFGESAGASVGMHL-----SPSRGLFHRAVLQSGA--P 263  
QY 260 SVPVDTYG-----NEIYDLFVSSAGC-----GSASDKLACLRASSSTLTDATNN 305  
Db 264 NGPW-ATVGMGEARRRATQLAHL-----VGPFGGTGGNDTELVAQLTRTPAQVLV---NH 315  
QY 306 TPGFLAYSSL-RLSYLPRPDGKNIITDDMYKLVRDGYASVVPVLIQDDEGTIF----- 358  
Db 316 EHWVLVQESVHRFSFVVPVVDGFLSTPEALINAGDFHGLQVLGVVVKDEGSFLVYGAP 375  
QY 359 GLSSLN--VTTNAQARAYFKQSFIIHASDAEIDTLMAAY-----PQD---ITQSPSPDTGV 408  
Db 376 GFSDKNESSLISRAEFLAGVRGVGPQVSDLAEEAVVLHYTDLHPEDPARLREALSDVVG 435  
QY 409 LNALTPQFKRISAVLGDIAFIHAR--RYFLNHPQG-----GTKYSLFKSLQSLGPI- 457  
Db 436 HNVVCP-----VAQLAGRLAAQGARVYAYVFEHRASTLSWPLMWGVPHGYEIEFIFGIP 491  
QY 458 -MGTTHANDIVWQDVLGSGSVIYNNAFIATATDLPNTA--GLLVNWPKYTSSSQ 510  
Db 492 PSRNTAEKIFAQRLM-----RYWANFARTGDPNEPRDKAPQWPPYTAGAQ 539  
  
RESULT 15  
PCT-US92-06106-2  
; Sequence 2, Application PC/TUS9206106  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Meir  
; TITLE OF INVENTION: ENZYMATICALLY ACTIVE RECOMBINANT HUMAN  
; TITLE OF INVENTION: ACETYLCHOLINESTERASE AND USES THEREOF  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John P. White, Esq.  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/06106  
; FILING DATE: 19920722  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 39304-B-PCT/JPW/EAB  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 664-0525  
; TELEX: 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 614 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US92-06106-2

Query Match 15.8%; Score 452; DB 5; Length 614;

```

Best Local Similarity 29.1%; Pred. No. 1.9e-35;
Matches 156; Conservative 83; Mismatches 179; Indels 118; Gaps 26;

Qy 35 AFLGIPPAEPVGNLRKDP---VPYSGSLNGQKFTSYGSPSCMQ-----QNPEGTFEEN 85
Db 62 AFLGIPPAEPVGNLRKDP---VPYSGSLNGQKFTSYGSPSCMQ-----QNPEGTFEEN 85
Qy 86 LKRTALDLVMSKVQFQAVLPQSECLATINVRPPTKAGANLPVMLTIFGGFEGISPTI 145
Db 119 PNREL-----SEDCLYLN-VMTYPYRPTSPVPLVWYIYGGFYSGASSL 161
Qy 146 PPPAQMVTKSYLMGKHIIH-----VAVNYVASWGLAGDDIKAGSGNAGLKDQRLGM 199
Db 162 -----DYDGRFLVQAERTVLVSNVYVGAFFLALPGSR-EAPGNVGLLDQRLAL 211
Qy 200 QWADNITAGFGDPSKVTIFGESAGSMVLCILWNGDNTYKGPPLFRAGIMOSGAMVP 259
Db 212 QWQENVAAGGDPSTVTLFGESAGAAVGMHLL-----SPPSRGLFHRVAVLQSGA--P 263
Qy 260 SDPVDGTYG-----NEIYDLFVSSAGC-----GSASDKLACLSASSDTLDDATNN 305
Db 264 NGPW-ATVGMGEARRATQLAHL-----VGCPPGCTGGNDTELVACLRTPRAQLV---NH 315
Qy 306 TPGFAYSSL-RLSYLPRPDGKNITDDMYKLVRDQKYASVPIIIGDQNDCTIF----- 358
Db 316 EWHVLPQESVFRFSFVPVWDGFLSDTPEALINAGDPHGLQVLVGVVWKEGSEYFLVYGAP 375
Qy 359 GLSSIN--VTTNAQARAYFKOSFIHASDAEIDTLMAY-----POD-----ITQGSFFDTGV 408
Db 376 GFSKDNESLISRAEFLAGVRVGVQVSDLAEEAVVLHYTDLWHPEDPARREALSDVVVGD 435
Qy 409 LNALTPOPKRISAVIGDLAFTHAR--RYFLNHFQ-----GTKYSFLSKQLSGLPI- 457
Db 436 HNVVCP-----VAQLAGRLAQAQARVYVFEHRASTLSWPLMGMVPHGYEIEFIEFGLD 491
Qy 458 -MGTFFHANDIYVQDYLLGSGSVIYNNAFIAFATDLDNTA--GLLVNMPKYTSSSQ 510
Db 492 PSRNYTAEKIFAQRLM-----RYWANFARTGDPNEPRDPKAPQWPPYTAGAQ 539

```

Search completed: July 29, 2004, 10:51:30  
Job time : 20 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 10:42:58 ; Search time 55 Seconds  
(without alignments)  
2810.063 Million cell updates/sec

Title: US-09-943-857-4  
Perfect score: 2864  
Sequence: 1 SMNSRGPAGRLGVSPTAKLA.....DNFRTAGYDALMTNPSFFV 547

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04: \*  
1: Geneseqp1980s: \*  
2: Geneseqp1990s: \*  
3: Geneseqp2000s: \*  
4: Geneseqp2001s: \*  
5: Geneseqp2002s: \*  
6: Geneseqp2003as: \*  
7: Geneseqp2003bs: \*  
8: Geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2864	100.0	547	6	ABU09071 Candida r
2	2588	90.4	547	6	ABU09074 Candida r
3	2495	87.1	547	6	ABU09073 Candida r
4	2459	85.9	534	4	AA847622 Synthetic
5	2434	85.0	547	6	ABU09072 Candida r
6	2396	83.7	547	6	ABU09070 Candida r
7	1107	38.7	563	2	AAR47577 Lipase of
8	1097	38.3	563	2	AAR10330 Gene prod
9	1078	37.6	557	2	AAR06370 Protein w
10	674.5	23.6	584	5	ABB07676 Carboxyle
11	685	23.2	575	5	ABB79538 Carboxyle
12	609	21.3	571	4	ABU53224 Human met
13	491.5	17.2	575	2	AAW39078 Torpedo c
14	488.5	17.1	575	2	AAW39079 Torpedo c
15	484	16.9	576	7	ABR84598 T califon
16	482	16.8	574	5	AAO18905 Rat butyr
17	477	16.7	574	5	AAO18903 Horse but
18	469	16.4	614	7	ADE61695 Rat Prote
19	469	16.4	614	7	ADE61687 Rat Prote
20	469	16.4	614	7	ADE61683 Rat Prote
21	469	16.4	614	7	ADE61691 Rat Prote
22	467	16.3	614	6	ABP59222 Human dru
23	462	16.1	826	2	AAV30100 Amino aci
24	462	16.1	826	2	AAV30101 An acetyl
25	459	16.0	574	5	AAO18904 Cat butyr

26	455	15.9	574	7	ABW00725 Human but
27	454	15.9	614	3	AAV49495 Human ace
28	454	15.9	614	3	AAV49494 Human ace
29	452	15.8	583	3	AG80773 AchE prot
30	452	15.8	600	2	AAW48797 Human ace
31	452	15.8	602	2	AAR37442 Full-leng
32	452	15.8	614	2	AAR80726 Human wil
33	452	15.8	614	3	AAV49489 Human wil
34	452	15.8	614	3	AAV49491 Human ace
35	452	15.8	614	5	AAU11231 Human ace
36	452	15.8	614	5	AAU11232 Human ace
37	452	15.8	614	5	AAU11234 Human ace
38	452	15.8	614	6	ABB99726 Amino aci
39	452	15.8	614	7	ADE61697 Human Pro
40	452	15.8	614	7	ADE61689 Human Pro
41	452	15.8	614	7	ADE61693 Human Pro
42	452	15.8	614	7	ADE61685 Human Pro
43	452	15.8	620	5	AAU11235 Human ace
44	451	15.7	574	5	AAE25235 Human but
45	451	15.7	574	5	AAO18898 Human but

## ALIGNMENTS

RESULT 1  
ABU09071  
ID ABU09071 standard; protein; 547 AA.  
XX

AC ABU09071;  
XX

15-JUL-2003 (first entry)  
XX

Candida rugosa lipase 3 polypeptide.  
XX

Lipase 3; enzyme; mutant; muten.  
XX

Candida rugosa.  
XX

Key Location/Qualifiers  
PH

Misc-difference 1..13  
FT /note= "Absent in wild-type Candida rugosa lipase"

Misc-difference 14  
FT /note= "Wild-type Ala substituted by Val"

Misc-difference 161  
FT /note= "Wild-type Pro substituted by His"

Misc-difference 408  
FT /note= "Wild-type Ile substituted by Val"

Misc-difference 409  
FT /note= "Wild-type Phe substituted by Leu"

Misc-difference 412  
FT /note= "Wild-type Ile substituted by Leu"

EP1288294-A2.  
XX

05-MAR-2003.  
XX

26-APR-2002; 2002EP-00009616.  
XX

31-AUG-2001; 2001US-00943857.  
XX

(SINI-) ACAD SINICA.  
XX

Tang S, Lee G, Shaw J;  
XX

WPI; 2003-395476/38.  
XX

N-PSDB; ABX95906.  
XX

Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for the preparation of Candida rugosa lipase for biocatalytic applications.

Claim 35; Page 5-7; 33pp; English.  
XX

CC The invention relates to an isolated mutant nucleic acid encoding a  
 CC Candida rugosa lipase polypeptide. The DNA has a sequence having at least  
 CC 80% identity to a wild-type DNA encoding Candida rugosa lipase. The  
 CC invention also relates to a microorganism comprising the DNA, where the  
 CC microorganism is a bacterium or yeast, preparing a mutant DNA encoding a  
 CC C.rugosa lipase and a chimeric C.rugosa lipase comprising a substrate  
 CC interacting domain of a first C.rugosa lipase and a non-substrate  
 CC interacting domain of a second C.rugosa lipase. The method is useful for  
 CC preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is  
 CC useful in the large scale manufacture of Candida rugosa lipase which is  
 CC useful for biocatalytic applications. This sequence represents Candida  
 CC rugosa lipase 3  
 XX  
 SQ Sequence 547 AA;

Query Match 100.0%; Score 2864; DB 6; Length 547;  
 Best Local Similarity 100.0%; Pred. No. 4,8e-245;  
 Matches 547; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 SMNSRGAGRLGVSPTAKLANGDITITGLNAINAEFLGIPPAEPVGNLRKDPVPYSGS 60  
 Db 1 SMNSRGAGRLGVSPTAKLANGDITITGLNAINAEFLGIPPAEPVGNLRKDPVPYSGS 60  
 Qy 61 LMGKFTSYGSCMQNPEGTFEENLKTALDLVMQSKVFOAVLPQSEDCLTINVRPPG 120  
 Db 61 LMGKFTSYGSCMQNPEGTFEENLKTALDLVMQSKVFOAVLPQSEDCLTINVRPPG 120  
 Qy 121 TKAGANLPVMLWIFGSGFEIGSPITFPPAQWTKSVLMGKHIIHVAVNYRVASGFLAGD 180  
 Db 121 TKAGANLPVMLWIFGSGFEIGSPITFPPAQWTKSVLMGKHIIHVAVNYRVASGFLAGD 180  
 Qy 181 DIKAGSGNAGLKDQRLQWQVADNIAGFGDPSKVITFBSAGSMVLCILINWDGNT 240  
 Db 181 DIKAGSGNAGLKDQRLQWQVADNIAGFGDPSKVITFBSAGSMVLCILINWDGNT 240  
 Qy 241 YKGKPLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSASDKLACLRSSDTLL 300  
 Db 241 YKGKPLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSASDKLACLRSSDTLL 300  
 Qy 301 DATNTTGFAYSLRLSYLPRDQKNTDDMYKLVRDGYASVPVIIGDNDGTTFGL 360  
 Db 301 DATNTTGFAYSLRLSYLPRDQKNTDDMYKLVRDGYASVPVIIGDNDGTTFGL 360  
 Qy 361 SSLNVTNQAQARAYFKQSFHASDAEIDTLMAYPQDITQGSPTDGVNALTQPKRIS 420  
 Db 361 SSLNVTNQAQARAYFKQSFHASDAEIDTLMAYPQDITQGSPTDGVNALTQPKRIS 420  
 Qy 421 AVLGDLPFIHARRYFLAHFQGGTKYSLSKQLSGLPIMGTTHANDIVWQDYLLGSGSVIY 480  
 Db 421 AVLGDLPFIHARRYFLAHFQGGTKYSLSKQLSGLPIMGTTHANDIVWQDYLLGSGSVIY 480  
 Qy 481 NNAEFAFATDLPNTAGLLVNWPKYTSQSSQGNLMNINALGLYTGKDNFTAGYDALMT 540  
 Db 481 NNAEFAFATDLPNTAGLLVNWPKYTSQSSQGNLMNINALGLYTGKDNFTAGYDALMT 540  
 Qy 541 NPSSFFV 547  
 Db 541 NPSSFFV 547

RESULT 2  
 ABU09074  
 ID ABU09074 standard; protein; 547 AA.  
 XX  
 AC ABU09074;  
 XX  
 DT 15-JUL-2003 (first entry)  
 XX  
 DE Candida rugosa lipase 8 polypeptide.  
 XX  
 KW Lipase 8; enzyme; mutant; mutein.  
 XX  
 OS Candida rugosa.

XX Key Location/Qualifiers  
 FH Misc-difference 1. 13  
 FT /note= "Absent in wild-type Candida rugosa lipase"  
 FT Misc-difference 14  
 FT /note= "Wild-type Ala substituted by Val"  
 FT Misc-difference 197  
 FT /note= "Wild-type Leu substituted by Met"  
 FT Misc-difference 266  
 FT /note= "Wild-type Ile substituted by Val"  
 FT Misc-difference 278  
 FT /note= "Wild-type Asn substituted by Asp"  
 FT Misc-difference 333  
 FT /note= "Wild-type Tyr substituted by Phe"  
 FT Misc-difference 343  
 FT /note= "Wild-type Asn substituted by Ser"  
 FT Misc-difference 344  
 FT /note= "Wild-type Ile substituted by Val"  
 FT Misc-difference 370  
 FT /note= "Wild-type Gln substituted by Glu"  
 FT Misc-difference 373  
 FT /note= "Wild-type Glu substituted by Gln"  
 FT Misc-difference 376  
 FT /note= "Wild-type Lys substituted by Thr"  
 FT Misc-difference 387  
 FT /note= "Wild-type Ile substituted by Leu"  
 FT Misc-difference 396  
 FT /note= "Wild-type Gly substituted by Gln"  
 FT Misc-difference 408  
 FT /note= "Wild-type Ile substituted by Val"  
 FT Misc-difference 427  
 FT /note= "Wild-type Gly substituted by Ala"  
 FT Misc-difference 429  
 FT /note= "Wild-type Thr substituted by Ile"  
 FT Misc-difference 430  
 FT /note= "Wild-type Leu substituted by His"  
 FT Misc-difference 530  
 FT /note= "Wild-type Phe substituted by Ser"  
 XX  
 EP1288294-A2.  
 XX  
 05-MAR-2003.  
 XX  
 26-APR-2002; 2002EP-00009616.  
 XX  
 31-AUG-2001; 2001US-00943857.  
 XX  
 (SINI-) ACAD SINICA.  
 XX  
 Tang S, Lee G, Shaw J;  
 XX  
 WPI; 2003-395476/38.  
 XX  
 N-PSDB; ABX95909.  
 XX  
 Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for  
 the preparation of Candida rugosa lipase for biocatalytic applications.  
 XX  
 Claim 35; Page 10-11; 33pp; English.  
 XX  
 The invention relates to an isolated mutant nucleic acid encoding a  
 CC Candida rugosa lipase polypeptide. The DNA has a sequence having at least  
 CC 80% identity to a wild-type DNA encoding Candida rugosa lipase. The  
 CC invention also relates to a microorganism comprising the DNA, where the  
 CC microorganism is a bacterium or yeast, preparing a mutant DNA encoding a  
 CC C.rugosa lipase and a chimeric C.rugosa lipase comprising a substrate  
 CC interacting domain of a first C.rugosa lipase and a non-substrate  
 CC interacting domain of a second C.rugosa lipase. The method is useful for  
 CC preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is  
 CC useful in the large scale manufacture of Candida rugosa lipase which is  
 CC useful for biocatalytic applications. This sequence represents Candida  
 CC rugosa lipase 8  
 XX  
 SQ Sequence 547 AA;

XX	Query Match	90.4%;	Score 2588;	DB 6;	Length 547;
PN	Best Local Similarity	88.7%;	Pred. No. 1.5e-220;		
XX	Matches 485;	Conservative 30;	Mismatches 32;	Indels 0;	Gaps 0;
PD					
XX					
XX	1	SNNSRGPAGRLGSVPTAKLANGDITGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGS	60		
PF					
XX	1	SNNSRGPAGRLGSVPTATLANGDITGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGS	60		
XX					
XX	61	LNQKFTSYGSCMQNPGTEENLGKTDALDVMQSKVFOAVLPQSEDLTINVRPFG	120		
XX					
XX	61	LDGQFTSYGSCMQNPGTEENLPKALDLMQSKVFEAVPSEDLTINVRPFG	120		
XX					
XX	121	TKAGANLPVLMWIFGGGFEIGSPTIPPPAQMYTKSVLMGKHIIHVAVNYRVASWGFLAGD	180		
DR					
XX	121	TKAGANLPVLMWIFGGGFEVGGTSTFPQAQMTKSTAMGKPIIHVSVNYRVSSWGFLAGD	180		
XX					
XX	181	DIKAEKSNAGLKDQRLGQWADNIAGFGDPSKVTIFGESAGSMVLCILWNGDNT	240		
XX					
XX	181	EIKAEKSNAGLKDQRMGMQWADNIAAFGGDPTKVTIFGESAGSMVLMCHILWNGDNT	240		
XX					
XX	241	YKGPFLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSASDKLACLSASDITLL	300		
XX					
XX	241	YKGPFLFRAGIMQSGAMVPSDAVDGYGNEIFDLASDAGCGSASDKLACLRGVSSDITLL	300		
XX					
XX	301	DATNTPGFLAYSRLSLYLPRPDGKNITDDMYKLVDRGKYASVPVLIIGDQNDGTFIPL	360		
XX					
XX	301	DATNTPGFLAYSRLSLYLPRPDGKNITDDMFALVREGKYASVPVLIIGDQNDGTFIPL	360		
XX					
XX	361	SSLNVTNAQARAYFKQSFHASDAEIDTLMAYPQDITQGSPPFTGVNALTTPQPKRIS	420		
XX					
XX	361	SSLNVTDAEARQYFTQSFVHASDAEIDTLMAYPQDITQGSPPFTGVNALTTPQPKRIS	420		
XX					
XX	421	AVLGLDLFIHARRYFLNHFQGGTKYSLKQSLGPIMGCTEHANDIVMODYLLGSGSVIY	480		
XX					
XX	421	AVLGLDLFIHARRYFLNHTYGGTKYSLKQSLGPIVGFTHANDIVMWHQFLGSGSVIY	480		
XX					
XX	481	NNAFTAFATDLPNTAGLLVNNPKYTSQSGNNLMINALGLYTKGNFRTAGYDALMT	540		
XX					
XX	541	NPSSFFV	547		
XX					
XX	541	NPPSFFV	547		
XX					
XX	RESULT 3				
XX	ABU09073				
XX	ID	ABU09073			
XX	AC	ABU09073			
XX	XX	ABU09073;			
XX	XX	15-JUL-2003 (first entry)			
XX	XX	Candida rugosa lipase 5 polypeptide.			
XX	DE	Lipase 5; enzyme; mutant; mutein.			
XX	XX	Candida rugosa.			
XX	XX				
XX	XX	Key	Location/Qualifiers		
XX	FT	Misc-difference 1. .13	/note= "Absent in wild-type Candida rugosa lipase"		
XX	FT	Misc-difference 14	/note= "Wild-type Ala substituted by Val"		
XX	FT	Misc-difference 160	/note= "Wild-type Lys substituted by Glu"		
XX	FT	Misc-difference 269	/note= "Wild-type Thr substituted by Ala"		
XX	FT	Misc-difference 359	/note= "Wild-type Gly substituted by Asp"		
XX	FT	Misc-difference 503	/note= "Wild-type Ser substituted by Tyr"		
XX	FT				

XX	EP1288294-A2.		
XX	05-MAR-2003.		
XX	26-APR-2002; 2002EP-00009616.		
XX	31-AUG-2001; 2001US-00943857.		
XX	(SINI-) ACAD SINICA.		
PA	Tang S, Lee G, Shaw J;		
XX	WPI; 2003-395476/38.		
DR	N-PSDB; ABX95908.		
XX	Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for the preparation of Candida rugosa lipase for biocatalytic applications.		
PT	Claim 35; Page 8-9; 33pp; English.		
XX	The invention relates to an isolated mutant nucleic acid encoding a Candida rugosa lipase polypeptide. The DNA has a sequence having at least 80% identity to a wild-type DNA encoding Candida rugosa lipase. The invention also relates to a microorganism comprising the DNA, where the microorganism is a bacterium or yeast, preparing a mutant DNA encoding a C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate interacting domain of a first C. rugosa lipase and a non-substrate interacting domain of a second C. rugosa lipase. The method is useful for preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is useful in the large scale manufacture of Candida rugosa lipase which is useful for biocatalytic applications. This sequence represents Candida rugosa lipase 5		
CC	Sequence 547 AA;		
XX			
XX	Query Match 87.1%; Score 2495; DB 6; Length 547;		
XX	Best Local Similarity 86.3%; Pred. No. 2.7e-212;		
XX	Matches 472; Conservative 31; Mismatches 44; Indels 0; Gaps 0;		
QY	1	SNNSRGPAGRLGSVPTAKLANGDITGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGS	60
DB	1	SNNSRGPAGRLGSVPTATLANGDITGLNAINAEFLGIPFAEPVGNLRFKDPVPYSGS	60
QY	61	LNQKFTSYGSCMQNPGTEENLGKTDALDVMQSKVFOAVLPQSEDLTINVRPFG	120
DB	61	LDGQFTSYGSCMQNPGTEENLPKALDLMQSKVFEAVPSEDLTINVRPFG	120
QY	121	TKAGANLPVLMWIFGGGFEIGSPTIPPPAQMYTKSVLMGKHIIHVAVNYRVASWGFLAGD	180
DB	121	TKAGANLPVLMWIFGGGFEVGGTSTFPQAQMYTKSVLMGKHIIHVAVNYRVASWGFLAGP	180
QY	181	DIKAEKSNAGLKDQRLGQWADNIAGFGDPSKVTIFGESAGSMVLCILWNGDNT	240
DB	181	EIKAEKSNAGLKDQRMGMQWADNIAAFGGDPTKVTIFGESAGSMVLMCHILWNGDNT	240
QY	241	YKGPFLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSASDKLACLSASDITLL	300
DB	241	YKGPFLFRAGIMQSGAMVPSDPVDTGYGNEIYDLFVSSAGCGSASDKLACLRGLSTQALL	300
QY	301	DATNTPGFLAYSRLSLYLPRPDGKNITDDMYKLVDRGKYASVPVLIIGDQNDGTFIPL	360
DB	301	DATNTPGFLAYSRLSLYLPRPDGKNITDDMYKLVDRGKYASVPVLIIGDQNDGTFIPL	360
QY	361	SSLNVTNAQARAYFKQSFHASDAEIDTLMAYPQDITQGSPPFTGVNALTTPQPKRIS	420
DB	361	SSLNVTDAEARAYLRKSFHASDAEIDTLMAYPQDITQGSPPFTGVNALTTPQPKRIS	420
QY	421	AVLGLDLFIHARRYFLNHFQGGTKYSLKQSLGPIMGCTEHANDIVMODYLLGSGSVIY	480
DB	421	AVLGLDLTFLSRRYFLNHTYGGTKYSLKQSLGPIVGFTHANDIVMWHQFLGSGSVIY	480
QY	481	NNAFTAFATDLPNTAGLLVNNPKYTSQSGNNLMINALGLYTKGNFRTAGYDALMT	540

```

|||||
481 NNAFIAFATDLDNPTAGLSVQWPKYTSSQAGDNLMQISALGLTYGKDNFRTAGYNALFA 540
|||||
541 NPSSFFV 547
:|:|:|
541 DPGHFFV 547

RESULT 4
AAB47622
ID AAB47622 standard; protein; 534 AA.
XX
AC AAB47622;
XX
DT 21-JAN-2002 (first entry)
XX
DE Synthetic lip1.
XX
KW Lipase; lip1; variant; ripening form; serine; 16-18C acyl chain.
XX
OS Candida rugosa.
XX
PN EP1130100-A1.
XX
PD 05-SEP-2001.
XX
PF 02-FEB-2001; 2001EP-00200375.
XX
PR 14-FEB-2000; 2000EP-00200513.
XX
PA (UNIL ) UNILEVER NV.
XX
PA (UNIL ) UNILEVER PLC.
XX
PI Brocca S, Bornscheuer UT, Pleiss J, Schmid RD, Schmid U;
PI Schmitt J;
XX
DR WPI; 2001-649825/75.
DR N-PSDB; AAB43625.
XX
PT Modified lipolytic enzymes with altered substrate specificity, useful for
PT biocatalytic applications comprising high specificity towards carbon 16
PT and carbon 18 acyl chains.
XX
PS Claim 1; Fig 1; 33pp; English.
XX
CC This sequence shows a parent lipase, lip1, derived from C. rugosa. The
CC lipase of the invention is a variant of this parent lipase, with altered
CC properties. The variant is the ripening form of C.rugosa lipase selected
CC from pre, pro, prepro or mature lipase, in which 60% or less of the CTG
CC codons encoding serine in the native C.rugosa sequence, are replaced by a
CC universal codon for serine. The modified nucleic acid sequence is further
CC modified, such that lipase variant exhibits an altered property. The
CC modified lipase is useful in a process requiring high specificity towards
CC 16-18C acyl chains
XX
SQ Sequence 534 AA;

Query Match 85.9%; Score 2459; DB 4; Length 534;
Best Local Similarity 86.7%; Fred. No. 4.1e-209;
Matches 462; Conservative 29; Mismatches 42; Indels 0; Gaps 0;

QY 15 PTAKLANGDTITGLNAINAEFLGIPFAEPVGNLRFKDPVPSGSLNGQKFTSYGSPSCM 74
DB 2 PTATLANGDTITGLNAINAEFLGIPFAEPVGNLRFKDPVPSGSLDQKFTSYGSPSCM 61

QY 75 QONPEGTEENLKGKTDLVNMQSKVQAVLQPSDECLTINVRPPGTAGANLPVMLWIF 134
DB 62 QONPEGTEENLKAALDLVNMQSKVFEAVSPSECLTINVRPPGTAGANLPVMLWTF 121

QY 135 GGGFEIGSPTTFPPAQMTKSVLMGKHIIHVAVNVRVASWGLAGDDIKARGSGNAGLKD 194
DB 122 GGGFETGTGTFPPAQMTKSIAMGKPIIHVSINVRVSSWGLAGDEIKARGSANAGLKD 181

```

```

195 QRLGMQWVADNIAGFGDPSKVTIFGESAGSMVLCILWNGDNTYKGPPLFRAGIMQS 254
182 QRLGMQWVADNIAAFGSDPTKVTIFGESTTSMVMCHILWNGDNTYKGPPLFRAGIMQS 241
255 GAWVPSPDVGTYGNEIYDLFVSSAGCGSASDKLACLRSSASDITLLDATNNTGFLAYSS 314
242 GAWVPSPDAVDGIYGNIEIFDLASNAGCGSASDKLACLRGVSSETLEDTATNNTPGFLAYSS 301
315 LRLSYLPRPDGKNITDDMYKLVDRDGKVASVPVLIIGDQNDDEGTIFGLSSLVNTTNAQARAY 374
302 LRLSYLPRPDGVITDDMYALVREGKYNIPVLIIGDQNDDEGTIFGLSSLVNTTDAQAREY 361
375 FKQSFHASDAEIDTLMAYPQDITQGSPEDTGVNLALTPOFKRISAVLGLAFIHARRY 434
362 FKQSFHASDTEIDTLMAYPQDITQGSPEDTGVNLALTPOFKRISAVLGLGFTLARRY 421
435 FLNHFQGGTKYSLSKQLSGIPMGTFPHANDIYWDYLLGSGSVIYNNAFAFATDLDPN 494
422 FLNHYTGGTKYSLSKQLSGIPVLGTGPHSNDIVFDYLLGSGSVIYNNAFAFATDLDPN 481
495 TAGLLVNWPKYTSSSQSGNNLMINLGLTYGKDNFRTAGYDALMTNPSSFFV 547
482 TAGLLVNWPEYTSSSQSGNNLMINLGLTYGKDNFRTAGYDALFENPPSFFV 534

RESULT 5
ABU09072
ID ABU09072 standard; protein; 547 AA.
XX
AC ABU09072;
XX
DT 15-JUL-2003 (first entry)
XX
DE Candida rugosa lipase 4 polypeptide.
XX
KW Lipase 4; enzyme; mutant; mutein.
XX
OS Candida rugosa.
XX
FH Key Location/Qualifiers
FT Misc-difference 1..13 "Absent in wild-type Candida rugosa lipase"
FT Misc-difference 14 "note= "Wild-type Ala substituted by Val"
XX
PN EP1288294-A2.
XX
PD 05-MAR-2003.
XX
PF 26-APR-2002; 2002EP-00009616.
XX
PR 31-AUG-2001; 2001US-00943857.
XX
PA (SINI-) ACAD SINICA.
XX
PI Tang S, Lee G, Shaw J;
XX
DR WPI; 2003-395476/38.
DR N-PSDB; ABX95907.
XX
PT Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for
PT the preparation of Candida rugosa lipase for biocatalytic applications.
XX
PS Claim 34; Page 7-8; 33pp; English.
XX
CC The invention relates to an isolated mutant nucleic acid encoding a
CC Candida rugosa lipase polypeptide. The DNA has a sequence having at
CC 80% identity to a wild-type DNA encoding Candida rugosa lipase. The
CC invention also relates to a microorganism comprising the DNA, where
CC microorganism is a bacterium or yeast, preparing a mutant DNA encoding a
CC C.rugosa lipase and a chimeric C.rugosa lipase comprising a substrate
CC interacting domain of a first C.rugosa lipase and a non-substrate
CC interacting domain of a second C.rugosa lipase. The method is useful for

```

CC preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is  
 CC useful in the large scale manufacture of Candida rugosa lipase which is  
 CC useful for biocatalytic applications. This sequence represents Candida  
 CC rugosa lipase 4  
 XX  
 SQ Sequence 547 AA;

Query Match 85.0%; Score 2434; DB 6; Length 547;  
 Best Local Similarity 83.5%; Pred. No. 7e-207;  
 Matches 457; Conservative 38; Mismatches 52; Indels 0; Gaps 0;

QY 1 SMNSRGPAGRLGVPVPAKLANGDTITGLNAINAEALGIPFAEPVGNLRFKDPVPYSGS 60  
 DB 1 SMNSRGPAGRLGVPVPAKLANGDTITGLNAINAEALGIPFAEPVGNLRFKDPVPYSGS 60  
 QY 61 LMGKFTSYGSPSCMQNPEGTPEENLGKLTALDLVMSQKVFQAVLPQSEDCLTINVRPPG 120  
 DB 61 LMGKFTSYGSPSCMQNPEGTPEENLGKLTALDLVMSQKVFQAVLPQSEDCLTINVRPPG 120  
 QY 121 TKAGANLPVMLWIFGGGFEIGSPITFPFAQMVTKSVLMGKHIIHVAVNYRVASWGLAGD 180  
 DB 121 TKAGANLPVMLWIFGGGFEIGSPITFPFAQMVTASVLMGKHIIHVAVNYRVASWGLAGD 180  
 QY 181 DIKAECSGNAGLKDQRLGMQWADNITAGFGGDPKVTIFGESAGSMVSLCHLIWNGDNT 240  
 DB 181 DIKAECSGNAGLKDQRLGMQWADNITAGFGGDPKVTIFGESAGSMVSLCHLIWNGDNT 240  
 QY 241 YGKPLFRAGIMQSGAMVSDPDVGYGNEIYDLFVSSAGCGSASDKLACLRASDSTLL 300  
 DB 241 YGKPLFRAGIMQSGAMVSDPDVGYGNEIYDLFVSSAGCGSASDKLACLRASDSTLL 300  
 QY 301 DATNTPGFLAYSSLRLSYLPRPDGKNTTDDMYKLVROGKYASVPVITGDQNDGTFGL 360  
 DB 301 QATSDTPGALAYSSLRLSYLPRPDGKNTTDDMYKLVROGKYASVPVITGDQNDGTFGL 360  
 QY 361 SSINVTYTAQARAYFKQSFIFHASDAEIDTLMAYPDITQGSFDTGVNALTPOFKRIS 420  
 DB 361 SSINVTYTAQARAYFKQSFIFHASDAEIDTLMAYPDITQGSFDTGVNALTPOFKRIS 420  
 QY 421 AVIGDLAFTHARRYFNLHFQGGTKYSFLSKQLSGLPIMGTFTHANDIVMDYLLGSGSVIY 480  
 DB 421 AVIGDLAFTHARRYFNLHFQGGTKYSFLSKQLSGLPIMGTFTHANDIVMDYLLGSGSVIY 480  
 QY 481 NNAFTAFATDLPNTAGLVNPKYTSQSGNNLMINALGLYTKGNFRTAGYDALMT 540  
 DB 481 NNAFTAFATDLPNTAGLVNPKYTSQSGNNLMINALGLYTKGNFRTAGYDALMT 540  
 QY 541 NPSSEFV 547  
 DB 541 NPSSEFV 547

RESULT 6  
 ABU09070  
 ID ABU09070 standard; protein; 547 AA.  
 XX  
 AC ABU09070;  
 XX  
 AC  
 XX  
 DT 15-JUL-2003 (first entry)  
 XX  
 DE Candida rugosa lipase 2 polypeptide.  
 XX  
 KW Lipase 2; enzyme; mutant; mutein.  
 XX  
 OS Candida rugosa.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1..13  
 FT /note= "Absent in wild-type Candida rugosa lipase"  
 FT Misc-difference 14  
 FT /note= "wild-type Ala substituted by Val"  
 FT Misc-difference 48  
 FT /note= "wild-type Thr substituted by Ser"

FT Misc-difference 91  
 FT /note= "Wild-type Arg substituted by Leu"  
 FT Misc-difference 92  
 FT /note= "Wild-type His substituted by Asp"  
 XX  
 PN EP1288294-A2.  
 PD  
 XX 05-MAR-2003.  
 XX  
 PF 26-APR-2002; 2002EP-00009616.  
 XX  
 PR 31-AUG-2001; 2001US-00943857.  
 XX  
 PA (SINI-) ACAD SINICA.  
 XX  
 PI Tang S, Lee G, Shaw J;  
 XX  
 DR WPI; 2003-395476/38.  
 DR N-PSDB; ABX95905.  
 XX  
 XX Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for  
 PT the preparation of Candida rugosa lipase for biocatalytic applications.  
 XX  
 PS Claim 35; Page 4-5; 33pp; English.  
 CC The invention relates to an isolated mutant nucleic acid encoding a  
 CC Candida rugosa lipase polypeptide. The DNA has a sequence having at least  
 CC 80% identity to a wild-type DNA encoding Candida rugosa lipase. The  
 CC invention also relates to a microorganism comprising the DNA, where the  
 CC microorganism is a bacterium or yeast, preparing a mutant DNA encoding a  
 CC C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate  
 CC interacting domain of a first C. rugosa lipase and a non-substrate  
 CC interacting domain of a second C. rugosa lipase. The method is useful for  
 CC preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is  
 CC useful in the large scale manufacture of Candida rugosa lipase which is  
 CC useful for biocatalytic applications. This sequence represents Candida  
 CC rugosa lipase 2  
 XX  
 SQ Sequence 547 AA;

Query Match 83.7%; Score 2396; DB 6; Length 547;  
 Best Local Similarity 82.3%; Pred. No. 1.7e-203;  
 Matches 450; Conservative 34; Mismatches 63; Indels 0; Gaps 0;

QY 1 SMNSRGPAGRLGVPVPAKLANGDTITGLNAINAEALGIPFAEPVGNLRFKDPVPYSGS 60  
 DB 1 SMNSRGPAGRLGVPVPAKLANGDTITGLNAINAEALGIPFAEPVGNLRFKDPVPYSGS 60  
 QY 61 LMGKFTSYGSPSCMQNPEGTPEENLGKLTALDLVMSQKVFQAVLPQSEDCLTINVRPPG 120  
 DB 61 LMGKFTSYGSPSCMQNPEGTPEENLGKLTALDLVMSQKVFQAVLPQSEDCLTINVRPPG 120  
 QY 121 TKAGANLPVMLWIFGGGFEIGSPITFPFAQMVTKSVLMGKHIIHVAVNYRVASWGLAGD 180  
 DB 121 TKAGANLPVMLWIFGGGFEIGSPITFPFAQMVTKSVLMGKHIIHVAVNYRVASWGLAGD 180  
 QY 181 DIKAECSGNAGLKDQRLGMQWADNITAGFGGDPKVTIFGESAGSMVSLCHLIWNGDNT 240  
 DB 181 DIKAECSGNAGLKDQRLGMQWADNITAGFGGDPKVTIFGESAGSMVSLCHLIWNGDNT 240  
 QY 241 YGKPLFRAGIMQSGAMVSDPDVGYGNEIYDLFVSSAGCGSASDKLACLRASDSTLL 300  
 DB 241 YGKPLFRAGIMQSGAMVSDPDVGYGNEIYDLFVSSAGCGSASDKLACLRASDSTLL 300  
 QY 301 DATNTPGFLAYSSLRLSYLPRPDGKNTTDDMYKLVROGKYASVPVITGDQNDGTFGL 360  
 DB 301 QATSDTPGALAYSSLRLSYLPRPDGKNTTDDMYKLVROGKYASVPVITGDQNDGTFGL 360  
 QY 361 SSINVTYTAQARAYFKQSFIFHASDAEIDTLMAYPDITQGSFDTGVNALTPOFKRIS 420  
 DB 361 SSINVTYTAQARAYFKQSFIFHASDAEIDTLMAYPDITQGSFDTGVNALTPOFKRIS 420  
 QY 421 AVIGDLAFTHARRYFNLHFQGGTKYSFLSKQLSGLPIMGTFTHANDIVMDYLLGSGSVIY 480

```

Db 421 ALLGDLAFTLARRYFLNYOGTKYSLKSLPLVLTGFHGNDDIINQDYLVGSGSVIY 480
QY 481 NNAPAFATDLDNPTAGLLVNWPKYTSOSSGNNLMNIALGLYTGKDNFRTRAGYDALMT 540
Db 481 NNAPAFATDLDNPKAGLWNWPTTSSOSSGNNLMQINGLGLYTGKDNFRPDAYSALFS 540
QY 541 NPSSEFFV 547
Db 541 NPSSEFFV 547

```

RESULT 7  
AAR47577  
ID AAR47577 standard; protein; 563 AA.

XX AAR47577;  
AC  
DT 16-OCT-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 19-JUL-1994 (first entry)

XX Lipase of Geotrichum candidum.

XX Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;  
KW Major cell wall protein; glycosyl-phosphatidyl-inositol;  
KW anchoring protein; alpha factor; alpha-agglutinin; invertase; inulinase;  
KW alpha-amylase; Geotrichum candidum; lipase; enzymatic process;  
KW fermentation; biodegradation; catalysis.

XX Galactomyces geotrichum.

XX WO9401567-A1.

XX 20-JAN-1994.

XX 07-JUL-1993; 93WO-EP001763.

XX 08-JUL-1992; 92EP-00202080.

XX 14-DEC-1992; 92EP-00203899.

XX (UNIL ) UNILEVER PLC.

XX (UNIL ) UNILEVER NV.

XX Kllis PM, Schreuder MP, Toschka H, Verrips CT;

XX WPI; 1994-035071/04.

XX N-ESDB; AAQ54020.

XX Immobilisation of enzymes to microbial cell wall - by prodn. of fusion  
XX protein of enzyme linked to anchoring protein.

XX Example 4; Page 49-52; 99pp; English.

XX The lipase is used in a method to immobilise enzymes to a microbial cell  
XX wall. The coding sequence is used in the production of a recombinant  
XX polynucleotide which comprises a structural gene encoding a protein with  
XX catalytic activity (the lipase) and at least part of a gene encoding at  
XX least the C-terminus of a protein capable of anchoring in a eukaryotic or  
XX prokaryotic cell wall. The anchoring fragment or protein is selected from  
XX alpha agglutinin, AGA 1, FLO 1, major cell wall protein of lower  
XX eukaryotes or a proteinase of lactic acid bacteria. The recombinant  
XX polynucleotide preferably also comprises a sequence encoding a signal  
XX peptide to ensure secretion of the expressed product. The signal peptide  
XX is preferably derived from glycosyl-phosphatidyl-inositol, anchoring  
XX protein, alpha factor, alpha-agglutinin, invertase or inulinase, alpha-  
XX amylase of Bacillus or proteinases of lactic acid bacteria. The host  
XX microorganism can be used for performing enzymatic processes on an  
XX industrial scale. (Updated on 25-MAR-2003 to correct PN field.) (Updated  
XX on 16-OCT-2003 to standardise OS field)

XX Sequence 563 AA;

```

Query Match 38.7%; Score 1107; DB 2; Length 563;
Best Local Similarity 43.4%; Pred. No. 4.4e-89;
Matches 240; Conservative 73; Mismatches 196; Indels 44; Gaps 9;

QY 9 GRIGSVPTAKLNGDITGLNALINEAFILGIPAEPPVGNLRFKDPVPYSGSLNGQKFTS 68
Db 16 GTLAQAFTAVLNGNEVISGLEKVDTFKGIPTADPEVGLRFXKHPQFFTGSYOGLKAND 75
QY 69 YGPMCOONPEGTFE-----ENLGTALDLVMSQKVQFQAVLPQSEDCLTINV 115
Db 76 FSSACMLQDFGNALFSLDKVGLGKILPDNLRLPYDMA-----QGSVSNEDCLYNV 129
QY 116 VRPPGTKAGANLPMWLIFGCGFEIGSPITPPPAQWMTKSVLMGKHIIHVAVNVRASWG 175
Db 130 FRPAGTKPDAKLPVMWVIYGGAFVFGSSASYPGNGYKESVEMQCPVVFVSVINTRTGPY 189
QY 176 FLAGDDIKAESGNAGLKDORLGMOWVADNIAGFGDPSKVTIFGESAGSMVSLCHLWN 235
Db 190 FLGGDAITAEAGTNAGLHDQKLEWSDNIANPGGDPDKMIFGESAGAMSVAHQLVAY 249
QY 236 DGNTYKKGKPLFRAGIMQSGAMVPSDPVDCGTGNEIYDLFVSSAGC---GSASDKLACLR 292
Db 250 GDNITYNGKQLFHSAILQSGGPLPYFDSTSVGPESAYSRTAQYAGCDTSASDNDTLACLR 309
QY 293 SASDITLLDATNN-----TPGFLAYSLSLSYLPRPDGKNITDDMYKLVRDGYAS 343
Db 310 SKSSDVLHSAQNSYDLKDLFLGLLPQFLGFG-----PRPDGNIIPDAAYELYSRYAK 362
QY 344 VPVIIGDNDGEGITFGLSLNVTNAQARAYFKOSFIHASDAEIDTILMAAYPQDITGSP 403
Db 363 VPYITGNOEDEGTILAPVAINATTPHVKWKLYICSQASDASLDRVLSLYPGWSGSP 422
QY 404 FDTGVNALTPQFKRISAVLGLDAFIHARRYFLNHFGGTYKYSFLSKQLSGL-PIMGTFH 462
Db 423 FRIGILNALTPQFKRIAAIFTDLLFQSPRRVMLNATKDVNRWTYATQLHNLVFLGTFH 482
QY 463 ANDIVWDYL-LGSGSVIYNNAFTAFATDLPNTAGLLVNWPKYTSOSSGNNLMNIAL 521
Db 483 GSDLLFYVVDLGPSSA-YRRYFISFANHHDPNVGTNLQOWDMYT---DAGKEMLQIHMI 538
QY 522 GLYTKDNFRTAG 534
Db 539 GNSMRTDDFRLEG 551

```

RESULT 8  
AAR10330  
ID AAR10330 standard; protein; 563 AA.

XX AAR10330;

XX 24-OCT-2003 (revised)

XX 25-MAR-2003 (revised)

XX 05-APR-1991 (first entry)

XX Gene product with lipase activity.

XX ATCC 34614.

XX Galactomyces geotrichum.

XX JP02299588-A.

XX 11-DEC-1990.

XX 27-MAR-1989; 89JP-00074721.

XX 27-MAR-1989; 89JP-00074721.

XX (KURK ) KURITA WATER IND LTD.

XX (OSAK ) OSAKA CITY.

XX WPI; 1991-027567/04.

DR N-PSDB; AAQ10313.

XX Gene for coding protein with lipase activity - is prepd. from messenger-  
PT ribonucleic acid of geo-trichum candidum ATCC 34614.

XX  
PT  
XX  
XX PS  
XX Claim 1; Fig 4; 12bp; Japanese.

XX The gene product may be isolated from a transformed expression sytem, and  
CC may be enhanced with stability in heat, alkali, acid and organic solvent  
CC by position-specific modulation. (Updated on 25-MAR-2003 to correct PA  
CC field.) (Updated on 24-OCT-2003 to standardise OS field)

XX  
SQ Sequence 563 AA;

Query Match 38.3%; Score 1097; DB 2; Length 563;  
Best Local Similarity 43.1%; Pred.No.3.4e-88;  
Matches 238; Conservative 72; Mismatches 200; Indels 42; Gaps 8;

QY 9 GRIGSVPTAKLANGDITITGLNAINAEPLGIPPAEPVGNLRFPKDPYPYSGSLNGQKQFTS 68  
Db 16 GTLAAQATVANGNEVLSGVLEGKVDYFKGIPFADPPVGLRFXKHPQPFSGYQGLKAND 75

QY 69 YGSPSCMOQNPEGTFE-----ENLKTALDLVMQSKVFQAVLPQSECLTINV 115  
Db 76 FSSACMQLDPGNAISLLDKVVGKGIIPDLNLRGLPYDMA-----QGSVMNEDCLVNV 129

QY 116 VRPPTGKAGANLPVWLIFGGPEIGSPTTFPPAQMVKSLMGKHLIHVAVNVRVASWG 175  
Db 130 FRPAGTKPDAKLPVWVIYGGAFVFGSSAYPGNAYVKESLEMGPVVVFVSINTRTGPYG 189

QY 176 FLAGDDIKAGSGNAGLKDQRLGQWVADNIAFGGDPSPKVTIFGESAGMSVLCHLIWN 235  
Db 190 FLGGDAITASEGNTNAGLHDKRGLEWSDNIANFEGGDPKVMIFGESAGMSVAHQLVAY 249

QY 236 DGNTYKXGKPLFRAGINQSGAMVPSDPVDGTGYGNEIYDLFVSSAGC--GSASDK--LACL 292  
Db 250 GGDNTYNGKQLFHSAILQSGGPIFYFDSTVGPESAYSREFAQYAGCDASAGDNETLACL 309

QY 293 SASSDTLLDATNN-----TPGFLAYSRLSYLPRDPGKNITDDMYKLVREDGKYAS 343  
Db 310 SKSSDVLHSQNSYDLKDLFGLLPQFLGFG-----PRDGNIPDAAYELRSRYAK 362

QY 344 VPVIQDNDEGTIFGLSSNVTNTNAQARAYFKOSFIHASDAEIDTLMAAYPQDITQSP 403  
Db 363 VPYITGNQDEGTILAPVAINATTPHVKKWLKYICSEASDASLDRVLSYPGSWSEGAP 422

QY 404 FDTGVLNALTPQKRISAVLGDLAFIHARVFLNHFGGTTKYSFLSKQLSGL--PIMGTEH 462  
Db 423 FRIGLINALTPQKRHAITFDLLQSPRRVMNATKDVNRWITYLATQLHNLFVFLGTEH 482

QY 463 ANDIVMQDYLKSGSVIYNNAFATADLDPNTAGLVNVPKYTSSSQSGNNLMMINALG 522  
Db 483 GSDLLFYAYAGPSSSAYRRYFISFANHDPNVAITNLKQMDMYT---DSGKEMLIQHIG 539

QY 523 LYTCKGNFRTAG 534  
Db 540 NSMRTDDFRIEG 551

RESULT 9  
AAR06370  
ID AAR06370 standard; protein; 557 AA.  
XX AC  
XX AAR06370;  
XX  
XX 24-OCT-2003 (revised)  
XX 25-MAR-2003 (revised)  
XX 13-DEC-1990 (first entry)  
XX  
XX Protein with lipase activity deduced from gene.  
XX  
XX Lipase.  
XX







AAW39078;  
 08-APR-1998 (first entry)  
 Torpedo californica acetylcholinesterase AChE.  
 Acetylcholinesterase; AChE; organophosphate; detoxification; oxime;  
 ray: chemical agent; treatment; prevention; aging; mutant.  
 Torpedo californica.  
 Key Location/Qualifiers  
 Modified-site 59..61  
 Region /note= "site as indicated in specification"  
 Region /note= "region as indicated in specification."  
 Region 254..265  
 Region /note= "region as indicated in specification"  
 Region 402..521  
 Modified-site /note= "region as indicated in specification"  
 Modified-site 416..418  
 Modified-site /note= "site as indicated in specification"  
 Modified-site 457..459  
 Modified-site /note= "site as indicated in specification"  
 Modified-site 569..571  
 Modified-site /note= "site as indicated in specification"  
 US5695750-A.  
 09-DEC-1997.  
 25-NOV-1994; 94US-00348920.  
 25-NOV-1994; 94US-00348920.  
 (USSA ) US SEC OF ARMY.  
 Radic Z, Taylor P, Maxwell D, Doctor BP, Saxena A;  
 WPI; 1998-041233/04.  
 Mutant Torpedo acetylcholinesterase - useful for detoxifying organo-  
 phosphate compounds.  
 Disclosure; Fig 1; 12pp; English.  
 This sequence represents an acetylcholinesterase (AChE) from Torpedo  
 californica which is used to create enzymes capable of detoxifying  
 organophosphates. Mutant versions of this AChE are useful when combined  
 with an oxime for detoxifying chemical warfare agents such as sarin or as  
 a means of treating or preventing the deleterious effects of  
 organophosphate exposure in organisms. Such variants can also destroy  
 organophosphates in an environmentally friendly manner and be more  
 resistant to aging than wild-type enzymes  
 Sequence 575 AA;  
 Query Match 17.2%; Score 491.5; DB 2; Length 575;  
 Best Local Similarity 30.4%; Pred. No. 1.9e-34;  
 Matches 164; Conservative 75; Mismatches 195; Indels 105; Gaps 24;  
 22 GDTITGLNAINAEFLGIPFAEPVGNLRFKDP---VPYSGSLNGQKFTSYGSCMQQNP 78  
 17 GTRVPVLSHSLS-AFLGIPFAEPVGNLRFKDP---VPYSGSLNGQKFTSYGSCMQQNP 71  
 79 EGTEENLGKTLALVMSQVQAVLPOSECLINV---VRPPGTGAGANLPVMLWIF 134  
 72 DEQPPFGSG-----SEMNPNREMSDCLVINIWPSPRPKST-----IVTWIY 116  
 135 GGGFEIGSPPTIPPAQWTKSLMGKH-----IIHVAVNYRVASWGFLA--GDDIKAE 186  
 117 GGGPYSGSSTL-----DYNGKYLAYTEEVVLVSLSYRVGAFGLAHGSG---EA 164

QY 187 SGNAGLKDQRLGMQWVADNIAGGDPKVTIIFGESAGSMVLCHLIWNGDNTYKGP 246  
 DB 165 PGNVGLLDQRMALQWHDNIQFEGDPKVTIIFGESAGSAGVGMHIL-----SPGSRDL 218  
 QY 247 FRAGIMQSGAMVPSDP---VDGTGNEIYDLFVSSAGCSASDK--LACLRSSSDTLDD 301  
 DB 219 FRRAILQSGS--PNCFWASVSAEGRRAVELGRNLNCLNLSDEELHLCREKKPQELID 276  
 QY 302 ATNNTPGFLAYSSL-RLSYLPRPDGKNITDDMKLVDRGKYASVPVLIIGQNDGDTIFGL 360  
 DB 277 VEWNV---LPFDSIFRFSFVPVIDGFFFTSLSEMLNSGNFKKTKQILLGVNKDEGSFLL 333  
 QY 361 -----SSLNVTYNAQAPAYFKQSFTHASDAEIDTLMAAYPQDITQGSFFDTGVNLAL 412  
 DB 334 YGAPGFSKDSKISREDFWGVKLSVPHANDLGLDAVTLOQ---TDMDDNNGIKNR- 388  
 QY 413 TPOFKRISAVLGDLPATHARRYFLN---HFQGTGKYSFLSKQISGL---PIMGTFHANDI 466  
 DB 389 ---DGLDDIVGDHNVICPLMHFNKVTYKFGNGTYLYFFNHRASNLVWPPEWGMVHGYEI 444  
 QY 467 VMODYLLGSGSVIYNN-----AFTAFATDLDLDPNTA-GLLVNWPKYTSSSQ 510  
 DB 445 ---EFVGLPLVKELVNTAEELSRRIHMYWATFAKTGNPNEPHSQESKWPLFTTKEQ 500  
 RESULT 14  
 AAW39079  
 ID AAW39079 standard; protein; 575 AA.  
 XX AC AAW39079;  
 XX DT 08-APR-1998 (first entry)  
 XX DE Torpedo californica acetylcholinesterase AChE mutant E199Q.  
 XX KW Acetylcholinesterase; AChE; organophosphate; detoxification; oxime;  
 XX ray: chemical agent; treatment; prevention; aging; mutant.  
 XX OS Torpedo californica.  
 XX Synthesis.  
 XX Key Location/Qualifiers  
 Modified-site 59..61  
 Region /note= "site as indicated in specification"  
 Region 67..94  
 Modified-site /note= "region as indicated in specification"  
 Region 199  
 Region /label= E199Q  
 Region /note= "wild-type Glu is replaced by Gln"  
 Region 254..265  
 Region /note= "region as given in specification"  
 Region 402..521  
 Region /note= "region as given in specification"  
 Modified-site 416..418  
 Modified-site /note= "site as given in specification"  
 Modified-site 457..459  
 Modified-site /note= "site as given in specification"  
 Modified-site 533..535  
 Modified-site /note= "site as given in specification"  
 US5695750-A.  
 09-DEC-1997.  
 25-NOV-1994; 94US-00348920.  
 25-NOV-1994; 94US-00348920.  
 (USSA ) US SEC OF ARMY.  
 Radic Z, Taylor P, Maxwell D, Doctor BP, Saxena A;  
 WPI; 1998-041233/04.

XX Mutant Torpedo acetylcholinesterase - useful for detoxifying organo-  
 PT phosphate compounds.  
 PS Claim 1; Fig 2; 12pp; English.  
 XX This sequence represents a variant of the enzyme acetylcholinesterase  
 CC (AChE) from Torpedo californica which is capable of detoxifying  
 CC organophosphates. This E199Q mutant has greater resistance to aging than  
 CC the wild-type AChE and can destroy organophosphates in an environmentally  
 CC friendly manner. When combined with an oxime, this mutant's ability to  
 CC detoxify several organophosphates is amplified and is predicted to have  
 CC applications for the detoxification of chemical warfare agents such as  
 CC sarin or as a means of treating or preventing the deleterious effects of  
 CC organophosphate exposure in organisms  
 XX Sequence 575 AA;  
 SQ  
 Query Match 17.1%; Score 488.5; DB 2; Length 575;  
 Best Local Similarity 30.2%; Pred. No. 3.6e-34;  
 Matches 163; Conservative 76; Mismatches 195; Indels 105; Gaps 24;  
 QY 22 GDTITGLNAINEAFGIPFAEPVGNLRFKDP---VPYSGSLNGQKFTSYGSPCMQNP 78  
 Db 17 GTRVPVLSHIS-AFGLGIPFAEPVGNMRFRPEPKPWSGVNASTY----PNNCQQYV 71  
 QY 79 EGTFEENLKTALDLVWQSKVQAVLPQSEDCITINV---VRPPGKTAGANLPVMLWTF 134  
 Db 72 DQFPFGSG-----SEMNPNREMSDCLYNIWVSPRPKST-----TVMWYI 116  
 QY 135 GGGPEIGSPITPPAQMTKSVLMGKH-----IIHVAVNRYVASWGFLA--GDDIKAE 186  
 Db 117 GGGFYSGSSTL-----DVYNGKYLAYTEEVVLVSLYRVGAFGLAHGSG---EA 164  
 QY 187 SQNAGLKQORLGMQWADNIAFGGDPKSVTIFFGESAGSMVLCHLIWNGDNTYKGP 246  
 Db 165 PGVUGLLDQRMALQWHDNIQFFGDKTITVIFGSGAGSAGVMHIL-----SPGSRDL 218  
 QY 247 FRAGIMQSGAMVPSDP---VDGTGNGEYDLFVSSAGCGSASDK--LACLRGASD 301  
 Db 219 FRRALQSGS--PNCFWASVVAEGRRAVELGRNLNLCNLSDEELHCLREKKPQELID 276  
 QY 302 ATNNTPGFLAYSSL-RLSYLPRPDGKNIITDDMYKLVDRGKYASVPVILIGDQND 360  
 Db 277 VEWNV---LPFDSIFRSPVVDGFEFTSLESMLNSGNFKKTOILLGVNKGDEG 333  
 QY 361 -----SSLVNTTNAQARAYFKOSFIHASDAEIDLTLMAAYPDITQSPDPTGV 412  
 Db 334 YGAPGFSKDSKISREDPMGKUSVPHANDGLDAVTLQY----TDMMDNNGIKNR- 388  
 QY 413 TPQFKRISAVLGLAFIHARRYFLN---HFQGGTKYFSLSKQLSGL---PIMGTFH 466  
 Db 389 ---DGLDDIVGDHNVICPLMHFVNKYTKFGNGTYLYFFNHRASNLVWPMWGV 444  
 QY 467 VQDYLGLSGSVIYNN-----AFIATDLDNPTA-GLLVNWKYTS 510  
 Db 445 ---EFVFGFLVPLKELNYTAEEALSRRIMHYWATEAKTGNPNPHSQQSKPLFT 500  
 RESULT 15  
 ABR84598  
 ID ABR84598 standard; protein; 576 AA.  
 XX  
 AC ABR84598;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE T californica acetylcholinesterase protein.  
 XX  
 KW Alpha-carboxylesterase; lipase; esterase; insect; biotransformation; ray;  
 KW biocatalysis; enzyme; hydrolysis; acetylcholinesterase.  
 XX  
 OS Torpedo californica.

XX WO2003066873-A1.  
 XX 14-AUG-2003.  
 XX 06-FEB-2002; 2002WO-AU000113.  
 XX 06-FEB-2002; 2002WO-AU000113.  
 XX (CSIR ) COMMONWEALTH SCI & IND RES ORG.  
 XX Oakeshott JG, Devonshire A, Coppin CW, Heidari R, Dorrian SJ;  
 PI Russell RJ;  
 XX WPI; 2003-767260/72.  
 XX Enzymatic reactions using an insect esterase or lipase, useful  
 PT particularly for resolution of hydrophobic insecticidal esters and  
 PT modification of fats and oils.  
 PT  
 PS Example 1; Page 77-79; 82pp; English.  
 XX The present invention relates to an enzyme-based biocatalytic process  
 CC where the enzyme is an insect esterase or lipase, or its mutant. Such an  
 CC enzyme may be an alpha-carboxylesterase. The method is used for  
 CC stereospecific hydrolysis of fatty acid and insecticidal (especially  
 CC pyrethrin) esters, forming optically pure acids or alcohols, for  
 CC modification of vegetable oils or fats for use in emulsions or other fat-  
 CC based foods, for preparing polymers, especially polyesters, for acylation  
 CC of substrates in leather tanning, in detergents, for removing pitch from  
 CC paper pulp, in biosensors (determination of triacylglycerols) and for the  
 CC preparation of regio- or chirally pure pharmaceutical or agrochemical  
 CC products. The present sequence is the ray acetylcholinesterase enzyme,  
 CC which was used in a homology comparison with a protein of the invention  
 XX  
 SQ Sequence 576 AA;  
 Query Match 16.9%; Score 484; DB 7; Length 576;  
 Best Local Similarity 30.6%; Pred. No. 9e-34;  
 Matches 160; Conservative 73; Mismatches 192; Indels 98; Gaps 23;  
 QY 35 AFLGIPFAEPVGNLRFKDP---VPYSGSLNGQKFTSYGSPCMQNPETFEENLKTAL 91  
 Db 30 AFLGIPFAEPVGNMRFRPEPKPWSGVNASTY----PNNCQQYVDVDFGFP 81  
 QY 92 DLVMSQKVFQAVLPQSEDCITINV-VRPPGKTAGANLPVMLWIFGGGFEIGSPITPPAQ 150  
 Db 82 -----SEMNPNREMSDCLYNIWVSPRPKSTAT---VMLWYGGFYSGSSTL----- 128  
 QY 151 MYTKSVLMGKH-----IIHVAVNRYVASWGFLA--GDDIKAEAGSAGLKQORLGMQW 202  
 Db 129 ---DVYNGKYLAYTEEVVLVSLYRVGAFGLAHGSG---EAPGNWGLLDQRMALQW 181  
 QY 203 ADNIAGFGDPKSVTIFFGESAGSMVLCHLIWNGDNTYKGPFRAGIMQSGAMVPSDP 262  
 Db 182 HDNIQFFGDKPTVTLFFGESAGRASVGMHIL-----SPGSRDLFRRALQSGS--PNC 233  
 QY 263 ---VDGTGNGEYDLFVSSAGCGSASDK--LACLRASDSTLLDATNTPGFLAYSSL-R 316  
 Db 234 WASVSAEGRRAVELRNLNLCNLSDEDLTQCLREKKPQELIDVEWNV---LPFDSIFR 290  
 QY 317 LSYLPRPDGKNIITDDMYKLVDRGKYASVPVILIGDQNDGTFGL-----SSLNVNTN 368  
 Db 291 FSVFVIDGFEFTSLESMLNAGNPKTKTOILLGVNKGDEGFFLYGAFGFSKDSKISR 350  
 QY 369 AQARAYFKOSFIHASDAEIDLTLMAAYPDITQSPDPTGVNLNTPQFKRISAVLGLDA 428  
 Db 351 EDFMSGVKLSVPHANDGLDAVTLQY----TDMMDNNGIKNR-----DGLDDIVGNHV 401  
 QY 429 IHARRYFLN---HFQGGTKYFSLSKQLSGL---PIMGTFHANDIVWQDYLGLSGSVIYNN 482  
 Db 402 ICPLMHFVNKYTKFGNGTYLYFFNHRASNLVWPMWGVHIGVEI---BFVFGFLVPLKELN 458

QY 483 -----AFIAFATDLPNTA-GLLVNWPXYTSSQ 510  
Db 459 YTAEEELSRIMHYWATFAKTGNENEPHSQESKWLPTTKEQ 501

Search completed: July 29, 2004, 10:49:10  
Job time : 58 secs



QY 75 QONPEGTFFE-----ENLGTALDLVMSQVFOAVLPQSEDCLTINVRPPT 121  
 DB 63 QLDGFGNAISLLDKVVGKILPDNLRLPLDMDMA-----QGSVSMNEDCLYLNVFRPAGT 116  
 QY 122 KAGANLPMWLWIFGGPGEIGSPITFPPAQMVTSLVLMGKHIHVAVNRYRVASWGLAGDD 181  
 DB 117 KPAKLPVWMIYGGAFVFGSSASYPGNGYKESVEMQPVVFSINRYTGPYGLGDA 176  
 QY 182 IKAEGSGNAGLKQORLGMOWADNIAGFGDPKSKVTIFGSGAGMSVYLCHLIWNGDNTY 241  
 DB 177 ITAEGSTNAGLHDKQKLEWVSDNIANFGGDPDKVMIFGSGAGMSVAHQVAYGGDNTY 236  
 QY 242 KGKPLFRAGIMQSGAMVPSDPVDTGYNEIYDLFVSSAGC---GSASDKLACLSASDDT 298  
 DB 237 NGKQLFHSAILQSGGPLPYFDSTSVGPESAYSRAQYAGCDTSASDNDTTLACLSKSDV 296  
 QY 299 LLDATNN-----TPGFLAYSLSRLSVLPDPGKNIITDDMKLVDRDGKIASVPVITG 349  
 DB 297 LHSQNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELYSRGYAKVPVITG 349  
 QY 350 QONDEGTIFGLSSLVNTTNAQARAYFKQSFHASDAEIDTLMAYPODITQGSFDTGVL 409  
 DB 350 NQDEGTILAPVAINATTPHVKKWLYICQASDASLDRLVSLYPGWSGSEGFRTGIL 409  
 QY 410 NALTPQKRIISAVLGDIAFHARRYFLNHFGGTYKYSFLSKLSGL-PIMGTTHANDIVM 468  
 DB 410 NALTPQKRIIAITFDLLFQSPRRVMLNATKDVNRWYTLATLQHLNLPFLGTFFHGSULD 469  
 QY 469 QDYL-LGSGSVIYNNAFIATFATDLPNTAGLLVNNPKYTSQSGNNLMINALGLYTK 527  
 DB 470 QYVVDLGPSSA-YRRYFISFANHDPNVGTLNQWDMYT---DAGKEMLIQHMVNSMRT 525  
 QY 528 DNFRTAG 534  
 DB 526 DDFRIEG 532

RESULT 2  
 Q00882 PRELIMINARY; PRT; 544 AA.  
 AC Q00882  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Lipase (Fragment).  
 OS Geotrichum candidum (Oospora lactis).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Dipodascaceae; Galactomyces.  
 OX NCBI\_TaxID=27317;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL Y-553;  
 RA Bertolini M., Laramée L., Thomas D., Cygler M., Schrag J., Vernet T.;  
 RT "Polymorphism in the lipase genes of Geotrichum candidum strains."  
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLSTERASE/LIPASE FAMILY.  
 DR EMBL; U02387; AAA03425.1; -.  
 DR PIR; S41093; S41093.  
 DR HSSP; P17573; 1THG.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR002018; Carbesterase.  
 DR InterPro; IPR000379; Ser\_estrs.  
 DR Pfam; PF00135; Coesterase; 1.  
 DR PROSITE; PS00122; CARBOXYLSTERASE\_B\_1; 1.  
 DR PROSITE; PS00941; CARBOXYLSTERASE\_B\_2; 1.  
 KW Hydrolase.  
 FT NON\_TER  
 SQ SEQUENCE 1 1 544 AA; 59264 MW; 96002E6652C23593 CRC64;  
 Query Match 38.0%; Score 1089; DB 3; Length 544;  
 Best Local Similarity 43.5%; Pred. No. 9,1e-68;  
 Matches 238; Conservative 73; Mismatches 192; Indels 44; Gaps 10;

QY 15 PTAKLANGDTITGLNAILNEAFLGIPPAEPVGNLRFKDPVPVYSGSLNGKQFTSYGSCM 74  
 DB 3 PTAIVNGNEVISGLVGGKVDTFKGIFFADPPVGLRFPKHPQPTGTSYQGLKANDFSSACM 62  
 QY 75 QONPEGTFFE-----ENLGTALDLVMSQVFOAVLPQSEDCLTINVRPPT 121  
 DB 63 QLDGFGNAISLLDKVVGKILPDNLRLPLDMDMA-----QGSVSMNEDCLYLNVFRPAGT 116  
 QY 122 KAGANLPMWLWIFGGPGEIGSPITFPPAQMVTSLVLMGKHIHVAVNRYRVASWGLAGDD 181  
 DB 117 KPAKLPVWMIYGGAFVFGSSASYPGNGYKESVEMQPVVFSINRYTGPYGLGDA 176  
 QY 182 IKAEGSGNAGLKQORLGMOWADNIAGFGDPKSKVTIFGSGAGMSVYLCHLIWNGDNTY 241  
 DB 177 ITAEGSTNAGLHDKQKLEWVSDNIANFGGDPDKVMIFGSGAGMSVAHQVAYGGDNTY 236  
 QY 242 KGKPLFRAGIMQSGAMVPSDPVDTGYNEIYDLFVSSAGC---GSASDKLACLSASDDT 298  
 DB 237 NGKQLFHSAILQSGGPLPYFDSTSVGPESAYSRAQYAGCDTSASDNDTTLACLSKSDV 296  
 QY 299 LLDATNN-----TPGFLAYSLSRLSVLPDPGKNIITDDMKLVDRDGKIASVPVITG 349  
 DB 297 LHSQNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELYSRGYAKVPVITG 349  
 QY 350 QONDEGTIFGLSSLVNTTNAQARAYFKQSFHASDAEIDTLMAYPODITQGSFDTGVL 409  
 DB 350 NQDEGTILAPVAINATTPHVKKWLYICSEASDASLDRLVSLYPGWSGSEGFRTGIL 409  
 QY 410 NALTPQKRIISAVLGDIAFHARRYFLNHFGGTYKYSFLSKLSGL-PIMGTTHANDIVM 468  
 DB 410 NALTPQKRIIAITFDLLFQSPRRVMLNATKDVNRWYTLATLQHLNLPFLGTFFHGSULD 469  
 QY 469 QDYL-LGSGSVIYNNAFIATFATDLPNTAGLLVNNPKYTSQSGNNLMINALGLYTK 527  
 DB 470 QYVVDLGPSSA-YRRYFISFANHDPNVGTLNQWDMYT---DAGKEMLIQHMVNSMRT 525  
 QY 528 DNFRTAG 534  
 DB 526 DDFRIEG 532  
 RESULT 3  
 Q00886 PRELIMINARY; PRT; 544 AA.  
 AC Q00886  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Lipase (Fragment).  
 OS Geotrichum candidum (Oospora lactis).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Dipodascaceae; Galactomyces.  
 OX NCBI\_TaxID=27317;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 34614;  
 RA Bertolini M., Laramée L., Thomas D., Cygler M., Schrag J., Vernet T.;  
 RT "Polymorphism in the lipase genes of Geotrichum candidum strains."  
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLSTERASE/LIPASE FAMILY.  
 DR EMBL; U02622; AAA03435.1; -.  
 DR PIR; S41090; S41090.  
 DR HSSP; P17573; 1THG.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR002018; Carbesterase.  
 DR InterPro; IPR000379; Ser\_estrs.  
 DR Pfam; PF00135; Coesterase; 1.  
 DR PROSITE; PS00122; CARBOXYLSTERASE\_B\_1; 1.  
 DR PROSITE; PS00941; CARBOXYLSTERASE\_B\_2; 1.  
 KW Hydrolase.  
 FT NON\_TER  
 SQ SEQUENCE 1 1 544 AA; 59263 MW; 97C96E7807DBF1EC CRC64;

```

Query Match      38.0%; Score 1088; DB 3; Length 544;
Best Local Similarity 43.5%; Pred. No. 1.1e-67;
Matches 238; Conservative 72; Mismatches 193; Indels 44; Gaps 10;

QY 15 PTAKLANGTTITGLNAINEAFGLIPFAEPVGNLRFKDPVYSGSLNGOKFTSYGPSM 74
D 3 PTAVLNGNEISGVLEKGVDTFKGIPFADPPVGLRFPKHPQPTFTGSGYQGLKANDFSACM 52
QY 75 QONPECTFE-----ENLKTALDVLQMSKVQAVLPQSEDCLTINVRPPT 121
D 63 QLDPGNAISLLDKVGLGIIPDLNGLPLYDMA-----QGSVMNEDCLYLNVFPAGT 116
QY 122 KAGANLPVMLTIFGGGFEIGSPITFPFAQMTKSVLMGKHIIHVAVNRYVASGFLAGDD 181
D 117 KPDAKLPVVMWYIGGAFVFGSSASYPGNGYVXESVEMGQPVVVFVSINRYTGPYGLGDA 176
QY 182 IKAEGSGNAGLDORLGMQWADNIAGFGDPSPKVTIFGESAGSMVSLCHLIWDGNTY 241
D 177 ITAEGNTNAGLHDQRKLEWSDNIANFGGDPDKVMIFGESAGMSVAHQLVAYGDN 236
QY 242 KGKPLFRAGIMOSGAMVPSDPDVTGTYGNEIYDLFVSSAGC-GSADK--LACLRSASDT 298
D 237 NGKQLFHSAILQSGPLPYFDSTVSGPESAYSRAQYAGCDASAGNETLACLRSSDV 296
QY 299 ILDATNN-----TPGFLAYSRLSYLPRPDGKNITDDMYKLVRDGVKASVPVITG 349
D 297 LHSAGNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELRYGRYAKVPYITG 349
QY 350 DONDEGTIFGLSLNVTTNAQARAYFKQSFHASDAEIDTLMAAYPQDITQSPPTGVL 409
D 350 NQDEGTILLAPVAINATTPHVKWLVKTCSEASDASLDRLVLSLXPGWSEGAFFRTGIL 409
QY 410 NALTPOFKRISAVLGLDLAFIARRYFLNHFGGTYKSFSLKSLGSL-PIMGTFHANDIYW 468
D 410 NALTPOFKRIAFITDLLFQSPRVMLNATKQVNRWYIATQHLNVLPFLGTFHGSDDLIF 469
QY 469 QDYL-LGSGSVYNNAFIATDLDNTAGLLVNNPKYTSOSSQSGNNLMINALGLYTK 527
D 470 QYVVDLGPSSA-YRRYFISFANHDPNVGTNLKQWDMYT---DSGKEMLIQHIMGNMST 525
QY 528 DNFRTAG 534
D 526 DDFRIEG 532

RESULT 4
ID P79066 PRELIMINARY; PRT; 563 AA.
AC P79066;
DT 01-MAY-1997 (TremBLrel. 03, Created)
DT 01-MAY-1997 (TremBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Lipase I precursor.
OS Geotrichum fermentans.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; mitosporic Dipodascaceae;
OC Geotrichum
OX NCBI_TaxID=44066;
RN [1]_TaxID=44066;
RP SEQUENCE FROM N.A.
RC STRAIN=WU-C12;
RA Arai T.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-188 FROM N.A.
RC STRAIN=WU-C12;
RX MEDLINE=97372566; PubMed=9228786;
RA Arai T., Yusa S., Kikimura K., Usami S.;
RT "Cloning and sequencing of the pcDNA encoding lipase I from
RT Trichosporon fermentans WU-C12."
RL FEMS Microbiol. Lett. 152:183-188 (1997).
CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.

DR EMBL; AB000260; BAA19072.1; -.
DR HSSP; P17573; 1THG.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR002018; CarboxylaseB.
DR InterPro; IPR000379; Ser_estrs.
DR Pfam; PF00135; Coesterase; 1.
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
KW Hydrolase; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 563 POTENTIAL.
SQ SEQUENCE 563 AA; 61548 MW; 210A12C206F33881 CRC64;

Query Match      38.0%; Score 1087; DB 3; Length 563;
Best Local Similarity 42.0%; Pred. No. 1.3e-67;
Matches 235; Conservative 83; Mismatches 200; Indels 42; Gaps 8;

QY 8 AGRLSVPTAKLANGTTITGLNAINEAFGLIPFAEPVGNLRFKDPVYSGSLNGOKFT 67
D 15 AGVLAQAPTAVLNGNEISGVLEKGVDTFKGIPFADPPVGLRFPKHPQPTFTGSGYQGLKAN 74
QY 68 SYGPSQMQONPECTFEENLKTALDVLW-QSKVF-----QAVLPOSEDCITIN 114
D 75 DFSPACMQJDPGNSL-----TLDDKALGLAKVPEEFGPLYDMAKGTVMNEDCLYN 128
QY 115 VVRPPTKAGANLPVMLTIFGGGFEIGSPITFPFAQMTKSVLMGKHIIHVAVNRYVASM 174
D 129 VVRPAGTKPDAKLPVVMWYIGGAFVYGSSAAYPGNSYVXESINMGQPVVVFVSINRYTGP 188
QY 175 GFLAGDDIKABGSGNAGLDORLGMQWADNIAGFGDPSPKVTIFGESAGSMVSLCHLIW 234
D 189 GFLGGDAITAEQNTWAGLHDQRKLEWSDNIANFGGDPDKVMIFGESAGMSVAHQLIA 248
QY 235 NQDNTYKGLPFRAGIMOSGAMVPSDPDVTGTYGNEIYDLFVSSAGC---GSADKLA 291
D 249 YGDMNTYKGLPFRAGIMOSGAMVPSDPDVTGTYGNEIYDLFVSSAGC---GSADKLA 291
QY 292 RGASDITLLDATNN-----TPGFLAYSRLSYLPRPDGKNITDDMYKLVRDGVK 342
D 309 RKSSSLVLDQAQSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELFSGRYA 361
QY 343 SVPTVIGQNDGTFIFGLSLNVTTNAQARAYFKQSFHASDAEIDTLMAAYPQDITQGS 402
D 362 KVPYISGNQDEGTAFAPVALNATTPHVKWLVQIFVDASEASIDRLVLSLPQTLVSGS 421
QY 403 PFDTCVLANALTPOFKRISAVLGLDLAFIARRYFLNHFGGTYKSFSLKSLGSL-PIMGTF 461
D 422 PRTGILNALTPOFKRVAAILSDMLFQSPRVMLSATKDVNRWYIATQHLNVLPFLGTF 481
QY 462 HANDIYWQDYLIGSGSVYNNAFIATDLDNTAGLLVNNPKYTSOSSQSGNNLMINAL 521
D 482 HGNELIFQFNWYNNIGPANSYLRVYFISFANHDPNVGTNLQWDMYT---DEGKEMLEIHT 538
QY 522 GLYTKDNFRAGYDALMTN 541
D 539 DNVMTDDYRIEGISNFET 558

RESULT 5
ID Q00883 PRELIMINARY; PRT; 544 AA.
AC Q00883;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Lipase (Fragment).
OS Geotrichum candidum (Oospora lactis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Galactomycetes.
OX NCBI_TaxID=27317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-552;

```

```

RA Bertolini M., Laramee L., Thomas D., Cygler M., Schrag J., Vernet T.;
RT "Polymorphism in the lipase genes of Geotrichum candidum strains.";
RC Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
DR EMBL; U02524; AAA03428.1; -.
DR PIR; S41092; S41092.
DR HSP; P17573; ITHG.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR002018; CarboxylesteraseB.
DR InterPro; IPR000379; Ser_estr.
DR Pfam; PF00135; Coesterase; 1.
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
DR Hydrolase.
KW NON TER
FT
SQ
SEQUENCE 544 AA; 59257 MW; 85DB41B1E28B8BFF1 CRC64;

Query Match 37.8%; Score 1084; DB 3; Length 544;
Best Local Similarity 43.3%; Pred. No. 2e-67;
Matches 237; Conservative 73; Mismatches 193; Indels 44; Gaps 10;

QY 15 PTKLANGDITITGLNAINAEFLGIPPAEPVGNLRKDPVPGSGSLNGOKFTSYGSPCM 74
DB 3 PTVANGNEVISGLVGKVDTFKGIIPADPVPVGLRFPKFPQPTGSGYQGLKANDFSSACM 62

QY 75 QONPEGTFE-----ENLGTALDLVMQSKVFOAVLPQSEDCLTINVRPPT 121
DB 63 QLDPGNAISLLDKVVGKGIIPDLNLRGLYDMA-----QGSVMNEDCLYLVNFRPAGT 116

QY 122 KAGANLPMVMIWIFGGFEGISPTIFFPPAQMTKSVLMKGIHIVAVNRYVASWGFLAGDD 181
DB 117 KPDAKLPVMWMIYGAFFVGSSASYPGNGYKESVEMGQPVVFSINVRTGPGYFLGGDA 176

QY 182 IKAGSGNAGLKQORLCQWVADNIAGFGDPSKVTIFGESAGSMVLCILHNDGNTY 241
DB 177 ITAEGNTNAGLHDQKGLWVSDNIANFGGDPDKVMIFGESAGAMSAHQILVAYGGDNTY 236

QY 242 KGKPLFRAGIMQSGAMVPSDPVDTGYNEIYDLVFSSAGC--GSASDK--LACLSASSDT 298
DB 237 NGKQLFHSAILQSGPLPYFDSTSVGPESAYSRFAQYAGCDASAGDNETLACLSKSDV 296

QY 299 LLDATNN-----TPGFLAYSSLRLSYLPRDGKNIITDDMYKLVRDGYASVPVIIG 349
DB 297 LHSQNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELYSRGYAKVPYITG 349

QY 350 DONDEGTFGLSSLVNTTNAQARAYFKQSFTHASDAEIDTLMAAYPQDITQGSFDTGVL 409
DB 350 NOEDEGTTILAPVAINATTPHVKWLYICSEASDAPLDRVLISLYPGSWSEGAPRTGIL 409

QY 410 NALTPQPKRIASVGLDLAFIHARRYFLNHFGGKTSPLSKQLSGL-PIMGTFHNDIVW 468
DB 410 NALTPQPKRIAAITDILLFQSPRRVLMNATKDVNRWTYLAQLHNLVPLFGLTFHGSLLF 469

QY 469 QDYL-LGSGSVIYNNAFATAFDLPNTAGLLVNWPKYTSQSGNNLMNINAGLYTGK 527
DB 470 QYYVDLGFSSA-YRRYFISFANHHPNVGTNLKQWDMYT---DAGKEMLIQIHMGNSMRT 525

QY 528 DNFRTAG 534
DB 526 DDFRIEG 532

RESULT 6
Q12614
AC Q12614; PRELIMINARY; PRT; 544 AA.
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Triacylglycerol lipase (EC 3.1.1.3) (Fragment).
OS Geotrichum candidum (Oospora lactis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Galactomycetes.

```

```

OX NCBI_TaxID=27317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 178.71;
RX MEDLINE=96049515; PubMed=7578238;
RA Phillips A., Pretorius G.H.J., Van Rensburg H.G.;
RT "Molecular characterization of a Galactomyces geotrichum lipase,
RL another member of the cholinesterase/lipase family.";
RL Biochim. Biophys. Acta 1252:305-311(1995).
CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
DR EMBL; X81656; CAAS7316.1; -.
DR PIR; S59958; S59958.
DR HSP; P17573; ITHG.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004806; F:triacylglycerol lipase activity; IEA.
DR InterPro; IPR002018; CarboxylesteraseB.
DR InterPro; IPR000379; Ser_estr.
DR Pfam; PF00135; Coesterase; 1.
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
DR Hydrolase.
KW NON TER
FT
SQ
SEQUENCE 544 AA; 59343 MW; 688072A3E43EEB39 CRC64;

Query Match 37.7%; Score 1081; DB 3; Length 544;
Best Local Similarity 43.1%; Pred. No. 3.3e-67;
Matches 236; Conservative 72; Mismatches 195; Indels 44; Gaps 9;

QY 15 PTKLANGDITITGLNAINAEFLGIPPAEPVGNLRKDPVPGSGSLNGOKFTSYGSPCM 74
DB 3 PTVANGNEVISGLVGKVDTFKGIIPADPVPVGLRFPKFPQPTGSGYQGLKANDFSSACM 62

QY 75 QONPEGTFE-----ENLGTALDLVMQSKVFOAVLPQSEDCLTINVRPPT 121
DB 63 QLDPGNAISLLDKVVGKGIIPDLNLRGLYDMA-----QGTVMSEDCLYLVNFRPAGT 116

QY 122 KAGANLPMVMIWIFGGFEGISPTIFFPPAQMTKSVLMKGIHIVAVNRYVASWGFLAGDD 181
DB 117 KPDKLPVMWMIYGAFFVGSSASYPGNGYKESVEMGQPVVFSINVRTGPGYFLGGDA 176

QY 182 IKAGSGNAGLKQORLCQWVADNIAGFGDPSKVTIFGESAGSMVLCILHNDGNTY 241
DB 177 ITAEGNTNAGLHDQKGLWVSDNIANFGGDPDKVMIFGESAGAMSAHQILVAYGGDNTY 236

QY 242 KGKPLFRAGIMQSGAMVPSDPVDTGYNEIYDLVFSSAGC---SASDKLACLSASSDT 298
DB 237 NGKQLFHSAILQSGPLPYFDSTSVGPESAYSRFAQYAGCDTSVSDNDTLACLSKSDV 296

QY 299 LLDATNN-----TPGFLAYSSLRLSYLPRDGKNIITDDMYKLVRDGYASVPVIIG 349
DB 297 LHSQNSYDLKDLFGLLPQFLGFG-----PRPDGNIIPDAAYELYSRGYAKVPYITG 349

QY 350 DONDEGTFGLSSLVNTTNAQARAYFKQSFTHASDAEIDTLMAAYPQDITQGSFDTGVL 409
DB 350 NOEDEGTTILAPVAINATTPHVKWLYICQASVASLDRVLISLYPGSWSEGAPRTGIL 409

QY 410 NALTPQPKRIASVGLDLAFIHARRYFLNHFGGKTSPLSKQLSGL-PIMGTFHNDIVW 468
DB 410 NALTPQPKRIAAITDILLFQSPRRVLMNATKDVNRWTYLAQLHNLVPLFGLTFHGSLLF 469

QY 469 QDYL-LGSGSVIYNNAFATAFDLPNTAGLLVNWPKYTSQSGNNLMNINAGLYTGK 527
DB 470 QYYVDLGFSSA-YRRYFISFANHHPNVGTNLQWDMYT---DAGKEMLIQIHMGNSMRT 525

QY 528 DNFRTAG 534
DB 526 DDFRIEG 532

RESULT 7
Q00892
ID Q00892 PRELIMINARY; PRT; 544 AA.
AC Q00892;

```



[illegible]



Db 350 NQDEGTAFAPVALNATTTTHVKKVQLQIFYDASEASIDRVLSLYPOTLSVGSFRTGIL 409  
 QY 410 NALTPQFKRISAVGLDAFIHARRYFLNHFQGGTKYSLKSLQSLG-L-PIMGTEHANDIWM 468  
 Db 410 NALTPQFKRAALSDMLFQSPRRVMSATKDVNRWYLSLTHLNLVPLFGTGHNELIF 469  
 QY 469 QYLLGSGSVIYNNAFIATDLPNTAGLLVNWPKYTSQSGNNIMINALGLYTGKD 528  
 Db 470 QFNVNITGPANSYLRYFISFANHDPNVGNTLLQWDQYT---DEGKEMLEIHMTDNVMRTD 526  
 QY 529 NPTAGYDALMTN 541  
 Db 527 DYRIEIGISNFETD 539  
 RESULT 11  
 Q8PN60 PRELIMINARY; PRT; 544 AA.  
 AC Q8PN60;  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Carboxylesterase, type B.  
 GN XAC1213.  
 OS Xanthomonas axonopodis (pv. citri).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OX NCBI\_TaxID=92829;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=306 / ATCC 13902 / XV 101;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Cannavan F., Cardoso J., Chambergro F., Ciapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.;  
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities";  
 RL Nature 417:459-463 (2002).  
 DR EMBL; AB011751; AAC36085.1; --  
 DR GO; GO:0005874; C:microtubule; IEA.  
 DR GO; GO:0004182; F:carboxypeptidase A activity; IEA.  
 DR GO; GO:0003824; F:catalytic activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0007018; F:microtubule-based movement; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR002453; Beta tubulin.  
 DR InterPro; IPR002018; Carboxylesterase.  
 DR InterPro; IPR000834; Peptidase\_M14.  
 DR InterPro; IPR000379; Ser. esters.  
 DR Pfam; PF00135; Coesterase; 1.  
 DR PROSITE; PS00122; CARBOXYLESTERASE\_B\_1; 1.  
 DR PROSITE; PS00941; CARBOXYLESTERASE\_B\_2; 1.  
 DR PROSITE; PS00133; CARBOXYPEPT\_ZN\_2; 1.  
 DR PROSITE; PS00228; TUBULIN\_B\_AUTOREG; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 544 AA; 57705 MW; 55CE24A079F17BFB CRC64;

Query Match 18.5%; Score 531; DB 16; Length 544;  
 Best Local Similarity 30.2%; Pred. No. 9.7e-29;  
 Matches 152; Conservative 70; Mismatches 192; Indels 90; Gaps 16;

Query Match 17.3%; Score 495; DB 13; Length 603;

QY 25 ITGLNAINIIEAFILGIPFAEPVGNLRFKDPVPYSGSLNGKFTSYGSCMQNPEGTFEE 84  
 Db 49 VRGVSAQGRAFLGVFPAAPPVUGELAFRAFPQPAATQVREDATQAGPACLPYRVG--- 104  
 QY 85 NLGKTALDLVMSQKVPQAVLPQSEDCLTINVRPPGTTKAGANLPVLMWIFGGFEGISGPT 144  
 Db 105 -----QKHV-----SEDCLTINVTAPPGPAPAHPRVAVMVWYIGGALELGSNV 146  
 QY 145 IEPFAQMTKSLVMGKHIIHVAVNYRVASWGFLAGDDIKAEAGSGNAGLKDORLGMQWVAD 204  
 Db 147 DYDLSALAAR-----QDVIVVAPNYRLGVFGFYAHFGLRGEGEGVALLDQQAALRWVOR 201  
 QY 205 NIAGFGGDPFKVTIFGESAGMSVLCHLIWNGDNTYKGPFRAGIMOSGAMVPSD--- 261  
 Db 202 NIAAFGGDAHNTVFGESAGANSI-CYQLASPG-----AAGLFQRAILQSGCLASDSSV 255  
 QY 262 FVDGTYGNEIYDLFVSSAGCGSASDKLACLRASASDTLIDATNTTGFPLAYSSLRSLYLP 321  
 Db 256 POREAESGGV--RMAQSLGCAHAPDAACLRALPADTLADAAFORRLGTGSD---AWAP 309  
 QY 322 RPDGKNITDDMYKLRDVGKVASVPVLIQDNDEGTIFG-----LSSLNVTNNAQARAYFK 376  
 Db 310 MSGGQALPLPPAAAIASGRHRVQPVLMGTRNDEGRULFAQLLSYIGKLNLRSGYEAV--- 366  
 QY 377 QSFIIHASDAEI-----DTLMAYPQDITQGSPPDTGVNLNALTQPKRISAVLGDLA 427  
 Db 367 -QRMHASPAVLRLQYAAVAQAQSRWEAFADIVTDGG-----FACPTRLGRAL 412  
 QY 428 FIHARRYFLNHFQGGTKYSLKSLQSLGPIMGTFHANDIWM---QDYLLGSGSVIYNNAF 484  
 Db 413 RTHAPVYAYEFDHPHAPYGLLRLPFS--PALGAFHASELVYLQRPWVL-SGRQFQSPAQ 469  
 QY 485 IAFATDL-----DPNTAG 497  
 Db 470 QAFANTLQDYWGAFARTGDPNGGG 493  
 RESULT 12  
 Q90ZK8 PRELIMINARY; PRT; 603 AA.  
 ID Q90ZK8;  
 AC Q90ZK8; (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Butyrylcholinesterase precursor (EC 3.1.1.8).  
 GN BCHE  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Geisler K., Chatonnet A., Layer P.G.;  
 RT "Chicken Butyrylcholinesterase";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.  
 DR EMBL; AJ306928; CAC37792.1; --  
 DR GO; GO:0004104; F:cholinesterase activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR002018; Carboxylesterase.  
 DR InterPro; IPR000997; Cholinesterase.  
 DR InterPro; IPR000379; Ser. esters.  
 DR Pfam; PF00135; Coesterase; 1.  
 DR PRINTS; PR00878; CHOLINESTRASE.  
 DR PROSITE; PS00122; CARBOXYLESTERASE\_B\_1; 1.  
 DR PROSITE; PS00941; CARBOXYLESTERASE\_B\_2; 1.  
 KW Hydrolase; Signal.  
 FT SIGNAL 1 29 POTENTIAL.  
 FT CHAIN 30 603 BUTYRYLCHOLINESTERASE.  
 SQ SEQUENCE 603 AA; 68480 MW; A350FDDF68574ADF CRC64;

Best Local Similarity 30.5%; Pred. No. 3.7e-26;  
Matches 163; Conservative 70; Mismatches 199; Indels 102; Gaps 22;

QY 22 GDTITGLNAINAFGIPFAEPVGNLRFKDPVPVSGSLNGQKFTSYGSCMQNDEGT 81  
DB 50 GGTVT-----AFLGIPYKGPGRGRFQKPEPEKWSGIWKATKANSCTY----- 95

QY 82 FEENLGTALDLVMQSKVFOAVLPQSEDCLTINVRP-PGKAGANLPVMLWIFGGGFEI 140  
DB 96 ----LIDTTPGFPFTEMMNPKTNLSDECLYNWIPSPKPK-----NATVWVWIIYGGSFET 148

QY 141 GSPTIFPPAQVTKSVLMGK-----HIHVAVNRYVASMGFLAGDDIKAEAGSNAGLKD 194  
DB 149 GSTSL-----PVYDGRFLARVERVIVVMNRYTGALGFLALPGNK-EVPGNAGLFD 198

QY 195 QRLGMQWADNIAGFGDPSKVTIFGSGAGSMVSLCHLWINDGNTYKGPLFRAGIMQS 254  
DB 199 QRLALQWQENIASFGGNPKSVTIFGSGAGSASVSYHIL-----SPKSHPLFTRAIMQS 252

QY 255 G-AMVPSDPVDGTGNEIYDLFVSSAGCGSA--SDKLACLRSSASDITLLDATNTPGFLA 311  
DB 253 GSANAPWAAITASARRRTVALAKQLKCPSTDETELILCLQDKPKDILE--NEVYVVKY 310

QY 312 YSSLRLSYLPRPDGKNITDDMYKLVRDGYKASVPVITIGDQNDGTIF-----GLS--SL 363  
DB 311 FSLHIIYFCPTVDGDFLADMEALIKNGIPKQTVLVGNKDEGTSFLVYGVPGFSKDS 370

QY 364 NVTNAQARAYFKOSFTHASDAEIDTLMAAY-----PQDITQSGPFDGTGLNALTP- 414  
DB 371 SLINKTOFEVALTISFPQVSKLAIESIIFQYTDWENQKPEHYRDAMDVDVIGDHIICPA 430

QY 415 -QFKRISAVLGLAFIHARRYFLNHFQGTGYKVSFLSKSLGSLPTMGTFPHANDIVQDYL 473  
DB 431 VEFKTAIEVGNVFF-----YFFE------RSSKLPWPEWGMHGEI---BEVF 474

QY 474 G---SGSVIYNA-----FIAPATLDLPNTAGLVN---WPKYTSSQ 510  
DB 475 GLPLERRVNYTKAEILSRSLRYWASPAKTGNPN--GTLINGTRWPVFTSTEQ 526

RESULT 13  
Q9XYA9  
ID O9XYA9 PRELIMINARY; PRT; 656 AA.  
AC O9XYA9  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Acetylcholinesterase (Fragment).  
GN ACHE.  
OS Meloidogyne javanica (Root-knot nematode).  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
OC Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.  
OX NCBI\_TaxID=6303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9270313; PubMed=10340488;  
RA Plette C., Arthaud L., Abad P., Rosso M.N.;  
RT "Molecular cloning of an acetylcholinesterase gene from the plant  
parasitic nematodes, Meloidogyne incognita and Meloidogyne javanica.";  
RL Mol. Biochem. Parasitol. 99:247-256(1999).  
CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.  
DR EMBL; AF080184; AAD25921.1; --  
DR HSP; P21836; 1MAA.  
DR GO; GO:0004104; F:cholinesterase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR002018; Carboxylesterase.  
DR InterPro; IPR000997; Cholinesterase.  
DR InterPro; IPR01064; Crystallin.  
DR InterPro; IPR000379; Sex esters.  
DR Pfam; PF00135; Coesterase; 1.  
DR PRINTS; PR00878; CHOLNESTRASE.  
DR PROSITE; PS00122; CARBOXYLESTERASE\_B\_1; 1.  
DR PROSITE; PS00941; CARBOXYLESTERASE\_B\_2; 1.  
DR PROSITE; PS00941; CARBOXYLESTERASE\_B\_2; 1.

DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
KW Hydrolase. 656 656  
FT NON TER 656 AA; AF90A221AEBFF00C CRC64;  
SQ SEQUENCE 656 AA; 76061 MW; 76061 MW; AF90A221AEBFF00C CRC64;

Query Match 16.9%; Score 484; DB 5; Length 656;  
Best Local Similarity 27.6%; Pred. No. 2.5e-25;  
Matches 156; Conservative 90; Mismatches 205; Indels 114; Gaps 23;

QY 19 LANGDITGLNAINA-----FLGIPFAEPVGNLRFKDPVPVSGSLNGQKFTSYG 71  
DB 49 LSDGSPIFG--SFQAATGKIYQFLGVPAEPPIGKLRFRPIKPEPWEQWNAITFRD 106  
QY 72 SCMQNDEGTFFENLGTALDLVMQSKVFOAVLPQSEDCLTINVRPPTGKAGANLPVL 131  
DB 107 SCV-QSPDTYFGFYGAT-----MNSNTPCEDCLYNIYVPEIDREKRLPVLF 156  
QY 132 WIFGGGFEIGSPTRIFPPAQVTKSVLMG-KHIIHVAVNRYVASMGFL-AGDDIKAEAGSN 189  
DB 157 WIYGGGFWSGTASL-----DVYDGKIFAGEENVIIIVTYRVTFVGFYLG---REAPGN 209  
QY 190 AGLKQDQRLGMQWADNIAGFGDPSKVTIFGSGAGSMVSLCHLWINDGNTYKGPLFRA 249  
DB 210 MGLWDQLLALKWYKNIQVFGDPSLITLFGESAGASVSMHML-----SPLSQPYFTR 263  
QY 250 GIMQSGAMVPSDPVDGTGYNE-----YDLFVSSAGCGSAS-----DKLAC 290  
DB 264 SILQSGAATAPWAVE-----NKQVALHRAVILYETM--KCGNGNMHSLAPDOWNMDEVLC 317  
QY 291 LRSASDTLLDATNTPGFLAYSSLR-----LSYLPRDGNKITDDMYKLVRDGYKVASVPV 346  
DB 318 LHAASAKLRDS-----EWSVMEFADFPPWVFVDGEFLVENIETSLKRGFKKTQL 369  
QY 347 IIGDNDGTIFGLSSLNVTNAQARAYFKOSFTHASDAEIDTLMAAYPQDITQSGPDT 406  
DB 370 LAGSNFDEAIYIVYQLS-DVFPFAEFKDKFKSRDVMKIVSSLLPRQILKSSIALQ 428  
QY 407 GVLNALTPQ-----FKRISAVLGLAF-IHARRYFLNHFQGT-----KYSFLSKQ 451  
DB 429 SILNYEPEGLPIESKNWVDSLDMGLDFLTCNVNEFALAHSEHGADTYTYMFESHASQ 488  
QY 452 LSGLPIMGTFHANDIVMVDYLLGSSSVIYNA-----FIAPATLDLP 493  
DB 489 QTPWPEWGMVGLHGEI---NFIFGP---YNRKQFKYKEQELSSRFMRFWANPARTSDP 542

494 N---TAGLLVNMWPKYTSSSQSGNNL 515  
543 NHNPDNSVISDPPYNSKTMVEYNL 567

RESULT 14  
O96529  
ID O96529 PRELIMINARY; PRT; 656 AA.  
AC O96529;  
DT 01-MAY-1999 (TRENBLrel. 10, Created)  
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Acetylcholinesterase.  
GN ACHE.  
OS Meloidogyne incognita (southern root-knot nematode).  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
OC Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.  
OX NCBI\_TaxID=6306;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9270313; PubMed=10340488;  
RA Plette C., Arthaud L., Abad P., Rosso M.N.;  
RT "Molecular cloning of an acetylcholinesterase gene from the plant  
parasitic nematodes, Meloidogyne incognita and Meloidogyne javanica.";  
RL Mol. Biochem. Parasitol. 99:247-256(1999).  
CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.  
DR EMBL; AF075718; AAD02835.1; --  
DR HSP; P21836; 1MAA.



Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2004, 14:51:20 ; Search time 4047 seconds  
(without alignments)  
16407.593 Million cell updates/sec

Title: US-09-943-857-3

Perfect score: 1532

Sequence: 1 gccaccgcgaagctcgcca.....tgaccacccgtcttcttg 1532

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.rod.\*

36: em.htg.man.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1532	100.0	1532	6	AX709921	AX709921 Sequence
2	1271.6	83.0	1855	8	CCLIP3	X66006 C.cylindrac
3	1191.6	77.8	1647	8	AY464467	AY464467 Candida c
4	1080.6	70.5	1541	6	AX709925	AX709925 Sequence
5	1025	66.9	1657	8	CCLIPASE	X16712 C. cylindra
6	1022	66.7	1733	6	A48377	A48377 Sequence 6
7	1022	66.7	1733	8	CCLIP1	X64703 C.cylindrac
8	1007.6	65.8	1511	6	AX709927	AX709927 Sequence
9	1002.4	65.4	1647	6	A81171	A81171 Sequence 1
10	998.4	65.2	1548	6	AX709923	AX709923 Sequence
11	982.4	64.1	1812	8	CCLIP5	X66007 C.cylindrac
12	937.2	61.2	1785	8	CCLIP4	X66007 C.cylindrac
13	867.6	56.6	2043	8	CCLIP2	X64704 C.cylindrac
14	830.4	54.2	1469	6	AX709919	AX709919 Sequence
15	831.8	42.5	1688	12	AF044078	AF044078 Synthetic
16	590.2	38.5	1605	6	A81172	A81172 Sequence 2
17	173.2	11.3	1635	8	GU02541	U02541 Geotrichum
18	173.2	11.3	1635	8	GU02625	U02625 Geotrichum
19	173	11.3	2045	6	A45569	A45569 Sequence 7
20	173	11.3	2045	6	A47607	A47607 Sequence 7
21	173	11.3	2045	6	AR019333	AR019333 Sequence
22	172.8	11.3	1635	8	GU02623	U02623 Geotrichum
23	171.2	11.2	1674	6	E02497	E02497 cDNA encodi
24	171.2	11.2	1767	8	GCALIP2	D0697 Geotrichum
25	171.2	11.2	1783	8	AB000260	AB000260 Geotrichu
26	164.2	10.7	1635	8	GCTAGL	X81556 G.candidum
27	164.2	10.7	1635	8	GU02525	U02525 Geotrichum
28	164.2	10.7	1828	6	A74255	A74255 Sequence 11
29	161.8	10.6	1635	6	A48376	A48376 Sequence 5
30	161.8	10.6	1635	8	GU02387	U02387 Geotrichum
31	159.4	10.4	1635	8	GU02622	U02622 Geotrichum
32	157.8	10.3	1692	6	E02678	E02678 cDNA encodi
33	156.2	10.2	1635	8	GU02524	U02524 Geotrichum
34	128.4	8.4	1934	8	AK107259	AK107259 Oryza sat
35	125.2	8.2	1692	8	GGECLR	X78032 G.geotrichu
36	120	7.8	4295	8	AY250996	AY250996 Aspergill
37	88.2	5.8	1716	6	AX268011	AX268011 Sequence
38	88.2	5.8	1738	6	AR034144	AR034144 Sequence
39	88.2	5.8	1738	6	BD005938	BD005938 Cephalosp
40	88.2	5.8	1738	8	AF025410	AF025410 Rhodospo
41	87	5.7	2000	8	ANGLA5	X56442 A.niger gla
42	70.8	4.6	13546	1	AE011751	AE011751 Xanthomon
43	70.2	4.6	6028	3	AF177382	AF177382 Culex tri
44	65.2	4.3	1520	3	CQSERES3	Z32696 C.quinquefa
45	65.2	4.3	1836	3	CQSERES1	Z32694 C.quinquefa

# ALIGNMENTS

RESULT 1	AX709921	1532 bp	DNA	linear	PAT 10-APR-2003
AX709921	Sequence 3 from Patent EP1288294.				
LOCUS	AX709921				
DEFINITION	Sequence 3 from Patent EP1288294.				
ACCESSION	AX709921				
VERSION	AX709921.1	GI:29786302			
KEYWORDS					
SOURCE	Candida rugosa				
ORGANISM	Candida rugosa				
REFERENCE	1				
AUTHORS	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
TITLE	Saccharomycetales; mitosporic Saccharomycetales; Candida.				
JOURNAL	Tang,S.J., Lee,G.C. and Shaw,J.F.				
	Recombinant candida rugosa lipases				
	Patent: EP 1288294-A 3 05-MAR-2003;				

Academia Sinica (TW)		Location/Qualifiers	
source		1..1532	
FEATURES		/organism="Candida rugosa"	
		/mol_type="unassigned DNA"	
		/db_xref="taxon:5481"	
ORIGIN			
Query Match 100.0%; Score 1532; DB 6; Length 1532;			
Best Local Similarity 100.0%; Pred. No. 6.1e-243;			
Matches 1532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	GCCACCGCCCAAGCTCGCAAGCGGACACCATCACCGGTCTCAACGCGCATCATCAACGA	60
DB	1	GCCACCGCCCAAGCTCGCAAGCGGACACCATCACCGGTCTCAACGCGCATCATCAACGA	60
QY	61	GGCGTTCTCGGCAATCCCTTTGGCGAGCCGCGGTGGGCAACCTTCGCTTCAAGGACCC	120
DB	61	GGCGTTCTCGGCAATCCCTTTGGCGAGCCGCGGTGGGCAACCTTCGCTTCAAGGACCC	120
QY	121	TGTGCGTACTTGGCTCGCTCAAGCGGCGAGAGTTACTTACGGCGCGTGCATGCAGCAG	180
DB	121	TGTGCGTACTTGGCTCGCTCAAGCGGCGAGAGTTACTTACGGCGCGTGCATGCAGCAG	180
QY	181	AACCCCGAGGCGACGTTTGAAGAGAACTTGGCAAGCGGCACTCGACTTGGTGATGCAG	240
DB	181	AACCCCGAGGCGACGTTTGAAGAGAACTTGGCAAGCGGCACTCGACTTGGTGATGCAG	240
QY	241	TCCAAGGTGTTCCAGCGCGTGTCTCCCGAGAGTGGACTGCCTCACCATCAACGTGGTG	300
DB	241	TCCAAGGTGTTCCAGCGCGTGTCTCCCGAGAGTGGACTGCCTCACCATCAACGTGGTG	300
QY	301	CGGCGCGCGGACCAAGCGGCGGCGCACTCCCGGTCTCGATCTTTGGCGGT	360
DB	301	CGGCGCGCGGACCAAGCGGCGGCGCACTCCCGGTCTCGATCTTTGGCGGT	360
QY	361	GGGTTTGAGATCGGAGCGCCACCATCTTCCCTCCCGCCAGATGGTCAACGAAGTGTG	420
DB	361	GGGTTTGAGATCGGAGCGCCACCATCTTCCCTCCCGCCAGATGGTCAACGAAGTGTG	420
QY	421	CTCATGGCAAGCCCATCATCACTGGCGCGTCAACTACCGTGTTCCTGGGGGTCT	480
DB	421	CTCATGGCAAGCCCATCATCACTGGCGCGTCAACTACCGTGTTCCTGGGGGTCT	480
QY	481	TGGCTGTGTACATCAAGCGCGGAGCGGCAAGCGCGGTGAGGACCAAGCGT	540
DB	481	TGGCTGTGTACATCAAGCGCGGAGCGGCAAGCGCGGTGAGGACCAAGCGT	540
QY	541	TGGGATGTCAGTGGGTGGCAGCAACATTTGCCGGGTTCGGCGCGACCCGAGCAAGGTGA	600
DB	541	TGGGATGTCAGTGGGTGGCAGCAACATTTGCCGGGTTCGGCGCGACCCGAGCAAGGTGA	600
QY	601	CATCTTTGGCAGGCGGCGAGCATGTCCGTGTGTGGCCACCTCATCTGGAACGAGCGGA	660
DB	601	CATCTTTGGCAGGCGGCGAGCATGTCCGTGTGTGGCCACCTCATCTGGAACGAGCGGA	660
QY	661	CAACAGTACAGGCGGCGGCGGCTTCCGCGCGGCATCATGCGGGAGCCATGTGTGC	720
DB	661	CAACAGTACAGGCGGCGGCGGCTTCCGCGCGGCATCATGCGGGAGCCATGTGTGC	720
QY	721	GGACCCGGTGGACGGCAGCTACGGCAACGAGATCTACGACCTCTTTGTCTCGAGTGTGG	780
DB	721	GGACCCGGTGGACGGCAGCTACGGCAACGAGATCTACGACCTCTTTGTCTCGAGTGTGG	780
QY	781	CTGTGGCAGCGCAGGCAAGCTCGGTGCTTGGCAGTGGGAGGACACCTTGTCTCA	840
DB	781	CTGTGGCAGCGCAGGCAAGCTCGGTGCTTGGCAGTGGGAGGACACCTTGTCTCA	840
QY	841	TGCCACCAACACATCTCCCTGGGTCTTGGCGTACTCTCTCTGGGTGTACTCCCGGCC	900
DB	841	TGCCACCAACACATCTCCCTGGGTCTTGGCGTACTCTCTCTGGGTGTACTCCCGGCC	900
QY	901	CGACGCGCAAGAACATCACCGATGACATGTAAGTTGGTGGCGGCGCAAGTATGCAAG	960
DB	901	CGACGCGCAAGAACATCACCGATGACATGTAAGTTGGTGGCGGCGCAAGTATGCAAG	960
QY	961	CGTTCCCGTGTATCTTTGGCGACGAGGAGGACCATCTTTGGCTCTTGAACGTG	1020
DB	961	CGTTCCCGTGTATCTTTGGCGACGAGGAGGACCATCTTTGGCTCTTGAACGTG	1020
QY	1021	ACCACGAATGCTCAGGCGCGGTCTTACTTCAAGCAGTTTCACTCCAGCGAGCGCGAG	1080
DB	1021	ACCACGAATGCTCAGGCGCGGTCTTACTTCAAGCAGTTTCACTCCAGCGAGCGCGAG	1080
QY	1081	ATCGACACCTTGTATGGCGCGTACCCCGAGGAGTCAACCCAGGCTCGGTTCGACGCTT	1140
DB	1081	ATCGACACCTTGTATGGCGCGTACCCCGAGGAGTCAACCCAGGCTCGGTTCGACGCTT	1140
QY	1141	CAACGCTCACCCCGCAGTTCAAGAGAACTCGCGGTCTCGGCGACCTTCATTCATCCACG	1200
DB	1141	CAACGCTCACCCCGCAGTTCAAGAGAACTCGCGGTCTCGGCGACCTTCATTCATCCACG	1200
QY	1201	CCCGCGCTACTTCTCAACCACTTCCAGGGCGGCAACCAAGTACTCTTCTCAAGCAGC	1260
DB	1201	CCCGCGCTACTTCTCAACCACTTCCAGGGCGGCAACCAAGTACTCTTCTCAAGCAGC	1260
QY	1261	TCGGGTTGCCAATCATGCGGACCTTCCATGCGCAACGACATTTGTGGCAGGACTACTTGT	1320
DB	1261	TCGGGTTGCCAATCATGCGGACCTTCCATGCGCAACGACATTTGTGGCAGGACTACTTGT	1320
QY	1321	TGGGAAGCGGAGCGGTCTTACAAACGCGGTTTATCGGTTCCGCCACCGACTTGGACC	1380
DB	1321	TGGGAAGCGGAGCGGTCTTACAAACGCGGTTTATCGGTTCCGCCACCGACTTGGACC	1380
QY	1381	CCAACACCGCGGGGTGTTGGTGAACCTGCCCAAGTACACGAGCAGCGGCAACAAC	1440
DB	1381	CCAACACCGCGGGGTGTTGGTGAACCTGCCCAAGTACACGAGCAGCGGCAACAAC	1440
QY	1441	TGATGATGATCAACGCTTGGGCTTGTATACCGGCAAGGACACTTCCGCCCGCTGGCT	1500
DB	1441	TGATGATGATCAACGCTTGGGCTTGTATACCGGCAAGGACACTTCCGCCCGCTGGCT	1500
QY	1501	ACGACGCGTTCATGACCAACCGCTTCTTTG 1532	
DB	1501	ACGACGCGTTCATGACCAACCGCTTCTTTTG 1532	
RESULT 2			
CCLIP3 1855 bp DNA linear PLN 30-JUN-1993			
LOCUS CCLIP3			
DEFINITION C.cylindracea LIP3 gene.			
ACCESSION X66006 S55937			
VERSION X66006.1 GI:296933			
KEYWORDS lip3 gene; lipase.			
SOURCE Candida cylindracea			
ORGANISM Candida cylindracea			
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
Saccharomycetales; mitosporic Saccharomycetales; Candida.			
REFERENCE 1 (bases 1 to 1855)			
AUTHORS Lotti,M., Grandori,R., Fusetti,F., Longhi,S., Brocca,S.,			
Tramontano,A. and Alberghina,L.			
Cloning and analysis of Candida cylindracea lipase sequences			
Gene 124 (1), 45-55 (1993)			
JOURNAL 93178975			
MEDLINE 840480			
PUBMED 840480			
REFERENCE 2 (bases 1 to 1855)			
AUTHORS Longhi,S.			
Direct Submission			
TITLE Submitted (06-MAY-1992) S. Longhi, Universita' degli Studi di			
Milano, Via Celoria 26, 20133 Milano, ITALY			
See X66006-8, X16712, X64703 & X64704			
COMMENT X66006, X66007 and X66008 are related sequences in the description			
of C.cylindracea Lip1 and Lip2 sequences (x64703, x64704).			
FEATURES			
Location/Qualifiers			
1..1855			
/organism="Candida cylindracea"			
source			



gene	/mol_type="genomic DNA"		
	/strain="ATCC 14830"		
	/db_xref="taxon:44322"		
	/clone_lib="SACI genomic"		
CDS	84..1733		
	/gene="LIP3"		
	84..1733		
	/gene="LIP3"		
	/EC_number="3.1.1.3"		
	/codon_start=1		
	/product="triacylglycerol lipase"		
	/protein_id="CAA46805.1"		
	/db_xref="GI:1325989"		
	/db_xref="GOA:P32947"		
	/db_xref="SWISS-PROT:P32947"		
	/translation="MKLALALLIASVAAPTAKLANGDTITGLNAINAEFLGIPFA EPVGNLRKPDVPVSGSLNGKFTLYGLPLCMQONPEGTFEENLGTALDLMQSKVF QAVLPSEDLTINVRPFTKAGANLPVLMWIFGGFEGISFTIPFPAMVTKSLVM KGPFIHVAVYVRVASFGLAGDDDKAEGSGNAGLKDORLGMQWADINAGPQDPSKV TIFELAGSMVSLCHLIMRSDNTYKRGKPLFRAGIMOLGAMVLPDVGTYGNEIYDL FVSSAGCSASVKLACLRALSADTLIDANTNPGLAYSLLRLLYLPDPDGKNI TDDM YKLVDRGKVASVPVITIGDNDSEGTIFGLLLNVTTNAQARAVFKOLF IHASDAEIDL MAAYPDITQGLPDTGIFNAITPPQKRIILAVLDLAFIHARRYELNHFQGTGKYSFL LKQLGLPMTGTFHNDIIVQDYLIGSGSVITINNAFIATDIDDPNTAGLLVNWPKYT SSQLGNLNMIMNALGLYTGKONFRTAGYDALMTNPLLEFV"		
sig_peptide	84..218		
	/gene="LIP3"		
mat_peptide	219..1730		
	/gene="LIP3"		
	/product="triacylglycerol lipase"		
	/EC_number="3.1.1.3"		
ORIGIN			
Query Match	83.0%;	Score 1271.6;	DB 8; Length 1855;
Best Local Similarity	94.9%;	Pred. No. 5.1e-200;	
Matches 1505;	Conservative 0;	Mismatches 19;	Indels 62; Gaps 16;
QY	2	CCCACGCCAAGCTCGCCACGCGGACACCATCACCGGTCTCAACGCCCATCATCAACGAG 61	
DB	132	CCCACGCCAAGCTCGCCACGCGGACACCATCACCGGTCTCAACGCCCATCATCAACGAG 191	
QY	62	GGTTCTCGGATTCCTTTTCGCGAGCGCGCGGTGGGAACTCTCGGTTCAGGACCT 121	
DB	192	GGTTCTCGGATTCCTTTTCGCGAGCGCGCGGTGGGAACTCTCGGTTCAGGACCT 251	
QY	122	GTGCGGTACTTGGCTCGCTCAACGCGCAGAGTTACTT-----ACGCGCGCTGCATG 174	
DB	252	GTGCGGTACTTGGCTCGCTCAACGCGCAGAGTTACTTGTACGGCGCGCTGTGCATG 311	
QY	175	CAGCAAAACCCGAGGCGCGCTTTGAAGAGAACCTTTGGCAAGACGCGACTCGACTTTGGTG 234	
DB	312	CAGCAAAACCCGAGGCGCGCTTTGAAGAGAACCTTTGGCAAGACGCGACTCGACTTTGGTG 371	
QY	235	ATGCAGTCAAGGTGTTCAGCGGTGCTTCCAGAGTAGAGACTGCGCTCACCATCAAC 294	
DB	372	ATGCAGTCAAGGTGTTCAGCGGTGCTTCCAGAGTAGAGACTGCGCTCACCATCAAC 431	
QY	295	GTGGTGGCGCGCGCGGACCAAGGCGGCGCAACTCCCGGTGATGCTCTGGATCTTT 354	
DB	432	GTGGTGGCGCGCGCGGACCAAGGCGGCGCAACTCCCGGTGATGCTCTGGATCTTT 491	
QY	355	GGCGTGGGTGTGAGATCGGCGCGCCCACTCTTCCCTCCCGCCAGATGTCACCAAG 414	
DB	492	GGCGTGGGTGTGAGATCGGCGCGCCCACTCTTCCCTCCCGCCAGATGTCACCAAG 551	
QY	415	AGTGTGCTATGGGCAAG--CCATCATCCAGTGGCGGTCAACTACCGGTGTGCTCGTG 473	
DB	552	AGTGTGCTATGGGCAAGCCATCATCCAGTGGCGGTCAACTACCGGTGTGCTCGTG 611	
QY	474	GGTTCTTGGCTGATGATCAAGCGCGGAGCGGCGGAAACCGCGGCTTGAAGGAC 533	
DB	612	GGTTCTTGGCTGATGATCAAGCGCGGAGCGGCGGAAACCGCGGCTTGAAGGAC 671	
QY	534	CHAGGTTTGGGATGCAAGTGGGTGGGAGAAACATTTCCCGGTTTCGGCGGACCCGAGC 593	
DB	672	CHAGGTTTGGGATGCAAGTGGGTGGGAGAAACATTTCCCGGTTTCGGCGGACCCGAGC 731	
QY	594	AAGGTGAC--ATCTTTGGGA---GGCGGCGAGCATGTCTGTTGTGGCCACCTCATCTGG 649	
DB	732	AAGGTGACATCTTTGGGAGCTGGCGGCGAGCATGTCTGTTGTGGCCACCTCATCTGG 791	
QY	650	AACGACGCGGCAACACAGTACAAAGGCAAGCGTGTCTCCGCGCGGCGCATCATGCA--G 706	
DB	792	AACGACGCGGCAACACAGTACAAAGGCAAGCGTGTCTCCGCGCGGCGCATCATGCAAGCTG 851	
QY	707	GGAGCCATGGTGC---GGACCCGCTGGACGCGACGTACGGCAACAGAGTCTACGACCTC 763	
DB	852	GGAGCCATGGTGC---GGACCCGCTGGACGCGACGTACGGCAACAGAGTCTACGACCTC 911	
QY	764	TTTGTCTCAGTGTCTGGCTGTGGCAGCGCGGAGCAAGCTCGCTGCTTGGCAGTGC-- 822	
DB	912	TTTGTCTCAGTGTCTGGCTGTGGCAGCGCGGAGCAAGCTCGCTGCTTGGCAGTGC-- 971	
QY	823	--GAGCGCACCTTGTCTCGATGCGCAACACACACCTCTCTGGGTTCTTGGCTACTCTCTCG 880	
DB	972	CTGAGCGCACCTTGTCTCGATGCGCAACACACCTCTCTGGGTTCTTGGCTACTCTCTCG 1031	
QY	881	TTGCGGTTGTACT-----CCCGGCGCGGACGCAAGAACATCACCATGACATGTACAAG 934	
DB	1032	TTGCGGTTGTACT-----CCCGGCGCGGACGCAAGAACATCACCATGACATGTACAAG 1091	
QY	935	TTGGTGGCGGACGGCAAGTATGCAAGCGTTCGGGTGATCATTTGGCGACAGACGAGAG 994	
DB	1092	TTGGTGGCGGACGGCAAGTATGCAAGCGTTCGGGTGATCATTTGGCGACAGACGAGAG 1151	
QY	995	GGCACCATCTTTG-----GCTCTTGAACGTGACCAAGTCTCAGGCGCGGTGCTTAC 1047	
DB	1152	GGCACCATCTTTGCTCTCTGCTGTGTGAACGTGACCAAGTCTCAGGCGCGGTGCTTAC 1211	
QY	1048	TTCAAGCA---GTTTCAACGCGGACGCGGAGATCGACACCTTGTATGGCGGCTAC 1104	
DB	1212	TTCAAGCAGCTGTTTCAATCCACGCGGAGATCGACACCTTGTATGGCGGCTAC 1271	
QY	1105	CCCGAGGACATCACCCAGGT---CGTTGACACG---TTCAACG---TCACCCCG 1154	
DB	1272	CCCGAGGACATCACCCAGGTCTGCGGTTGACACCGGATCTTCAACGCAATCACCCCG 1331	
QY	1155	CAGTTCAAGAGAAATC---GCGGTGCTCGCGACCTTGCATTCATCCACGCGCGGCTAC 1211	
DB	1332	CAGTTCAAGAGAAATCCTGCGGTGCTCGCGGACCTTGCATTCATCCACGCGCGGCTAC 1391	
QY	1212	TTCTCAACCACTTCCAGGCGGACCAAGTACTGTTCTCTC-----AAGCAGTCTCGG 1265	
DB	1392	TTCTCAACCACTTCCAGGCGGACCAAGTACTGTTCTCTC-----AAGCAGTCTCGG 1451	
QY	1266	TTGCCNATCATGGGACCTTCCATGCGCAAGCATTTGTGGCAGGACTACTTGTGGGA 1325	
DB	1452	TTGCCNATCATGGGACCTTCCATGCGCAAGCATTTGTGGCAGGACTACTTGTGGGA 1511	
QY	1326	AGCGGCGGCTCATCTTACAAACAAAGCGTTTATCGGTTCCGACCACTTTGACCCCAAC 1385	
DB	1512	AGCGGCGGCTCATCTTACAAACAAAGCGTTTATCGGTTCCGACCACTTTGACCCCAAC 1571	
QY	1386	ACCGGGGGTGTGTGGTGAACCTGGCCCAAGTACACGAGCAGC-----CAGGGCAACAAC 1439	
DB	1572	ACCGGGGGTGTGTGGTGAACCTGGCCCAAGTACACGAGCAGC-----CAGGGCAACAAC 1631	
QY	1440	TTGATGATGATCAACGCTTGGGCTTTGATACCGGCAAGGACAACTTCCGACCGGTGCG 1499	
DB	1632	TTGATGATGATCAACGCTTGGGCTTTGATACCGGCAAGGACAACTTCCGACCGGTGCG 1691	
QY	1500	TACGACGCTTGTATGACCAACCCGTT 1525	
DB	1692	TACGACGCTTGTATGACCAACCCGTT 1717	

```

RESULT 3
AY464467      1647 bp      mRNA      linear      PLN 03-DEC-2003
LOCUS        Candida cylindracea lipase mRNA, partial cds.
DEFINITION
ACCESSION    AY464467
VERSION      AY464467.1  GI:38565535
KEYWORDS
SOURCE
ORGANISM     Candida cylindracea
              Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
              Saccharomycetales; Mitosporic Saccharomycetales; Candida.
REFERENCE
  1 (bases 1 to 1647)
  Wang J.-Z., Wei D.-Z. and Zhang L.-Y.
  Cloning and sequence analysis of a lipase gene from Candida rugosa
  Unpublished
  2 (bases 1 to 1647)
  Wang J.-Z., Wei D.-Z. and Zhang L.-Y.
  Direct Submission
  TITLE
  JOURNAL
  AUTHORS
  TITLE
  JOURNAL
  Biochemistry, Meilong Road, Shanghai 200237, P.R. China
  Location/Qualifiers
  FEATURES
    source      1..1647
                /organism="Candida cylindracea"
                /mol_type="mRNA"
                /db_xref="taxon:44322"
    CDS         1..1647
                /BC_number="3.1.1.3"
                /codon_start=1
                /product="lipase"
                /protein_id="AAR24090.1"
                /db_xref="GI:38565536"
                /translation="MELALLLIASVAAPAKLANGDTITGLNAINBAPLIPFA
                EPPVNLRFKDPVPYSGSLNGKFTLYCPLMQONPEGTPEENIGKTLADLMVQSKVP
                QAVLPQSEDLITINVRPPGTHKAGANLPVLMWIFGGFEIGSPITFPPAQWTKSLM
                KPLIHVAVNTRVSGFLAGDDIKABSGNAGLQDRLGMQWADNINAGPQDFSKV
                TIFELAGSLSVLCHLIWNGDNTYKGLPFRAGIMQIGAMVPLDPVDYGVNIEYDL
                FVSSAGCGSADKLALRSALSDLLDNTNTPGLAYSLRLLYLPRPDGKNITIDM
                YKLVRDGKASVPVIGQNDQGTFFGLINLVTTDAEAOYFKELFVHTSDABLDTL
                MTAYPDQITQSPDGTGLNALTPQKRLILAVLGLDAFIHARRVFLNHYTGTVKVSFL
                LKQLLGLAVGTGFHNDIVFDYLVGSGSVIYNNAFIATVTDLPNTAGLLVNWFEYT
                SSULQGNLMMINALGLYTGKDNFRTAGIDALMANPPLFFV"
    ORIGIN
      Query Match      77.8%; Score 1191.6; DB 8; Length 1647;
      Best Local Similarity 91.4%; Pred. No. 8.1e-187;
      Matches 1448; Conservative 0; Mismatches 74; Indels 62; Gaps 15;
      QY      2   CCCACGCCAAGCTCGGCAACGGCGACACCATCAACCGGTCTCAACGCCATCATCAACGAG 61
      DB      49   CCCACGCCAAGCTCGGCAACGGCGACACCATCAACCGGTCTCAACGCCATCATCAACGAG 108
      QY      62   GCGTTCCTCGCATTCCTTTCCGAGCGCGCGGTGGCGACCTCCGCTTCAAGGACCCCT 121
      DB      109   GCGTTCCTCGCATTCCTTTCCGAGCGCGCGGTGGCGACCTCCGCTTCAAGGACCCCT 168
      QY      122   GTGCGGTACTCTGGTCTCGCTCAACGGCGCAGAGTACTT-----ACGGCGCGGTGCATG 174
      DB      169   GTGCGGTACTCTGGTCTCGCTCAACGGCGCAGAGTACTTCTCTAGCGGCGCGGTGTGCATG 228
      QY      175   CAGCAGAACCCCGAGGGCACTTTGAAGAGAACCTTTGGCAGAACGGCAGCTCGACTGTGGTG 234
      DB      229   CAGCAGAACCCCGAGGGCACTTTGAAGAGAACCTTTGGCAGAACGGCAGCTCGACTGTGGTG 288
      QY      235   ATGCAGTCCAGGTGTTCCAGCGGTGCTTCCAGAGTGAGGACTGCTCCACCATCAAC 294
      DB      289   ATGCAGTCCAGGTGTTCCAGCGGTGCTTCCAGAGTGAGGACTGCTCCACCATCAAC 348
      QY      295   GTGGTGGCGCGCGGGCACCAAGCGCGCGCAACCTCCCGGTGATGCTCTGGATCTTT 354
      DB      349   GTGGTGGCGCGCGGGCACCAAGCGCGCGCAACCTCCCGGTGATGCTCTGGATCTTT 408
      QY      355   GCGCGTGGGTTTGAAGATCGGACGCCCAACATCTTCCCTCCCGCCAGATGTTACCAAG 414

```





Db 1396 AGCGGCTCGCTCATCTACAAACACGCGTTCATTGCGCTTTGCCACGACTTGGACCCCAAC 1455  
Qy 1386 ACCGGGGGTGTGTGTAAGTGGCCCAAGTACACGAGCAGC-----CAGGGCAACAC 1439  
Db 1456 ACCGGGGGTGTGTGTAAGTGGCCCGAGTACACGAGCAGCTGAGCTGGGCAACAC 1515  
Qy 1440 TTGATGATGATCAACCGCTTTGGGCTTTGTACACCGGCAAGGACAACTTCGCGCACCGTGGC 1499  
Db 1516 TTGATGATGATCAACCGCTTTGGGCTTTGTACACCGGCAAGGACAACTTCGCGCACCGCGC 1575  
Qy 1500 TAGGACCGCTTGTATGACCAACCG 1523  
Db 1576 TAGGACCGCTTGTCTCCAACCG 1599

RESULT 6  
A48377  
LOCUS A48377 1733 bp DNA linear PAT 07-MAR-1997  
DEFINITION Sequence 6 from Patent WO9603511.  
ACCESSION A48377  
VERSION A48377.1 GI:2302168  
KEYWORDS  
SOURCE Candida rugosa  
ORGANISM Candida rugosa  
Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; mitosporic Saccharomycetales; Candida.  
REFERENCE 1 (bases 1 to 1733)  
AUTHORS Alibert G., Mouloungui, Z. and Boudet, A.  
TITLE METHOD FOR PRODUCING FATTY ACIDS OR DERIVATIVES THEREOF FROM OIL  
PLANT  
JOURNAL Patent: WO 9603511-A 6 08-FEB-1996;  
TOULOUSE INST NAT POLYTECH (FR)  
COMMENT Other publication AU 2984995 960222  
Other publication FR 2722798 960126.  
FEATURES  
source Location/Qualifiers  
1..1733  
/organism="Candida rugosa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:5481"

ORIGIN  
Query Match 66.7%; Score 1022; DB 6; Length 1733;  
Best Local Similarity 84.7%; Pred. No. 7.4e-159;  
Matches 1342; Conservative 0; Mismatches 180; Indels 62; Gaps 15;  
Qy 2 CCACCGCAAGCTCGCCACGGGACACCATCACCGTCTCAACGCCATCATCAACGAG 61  
Db 49 CCACCGCACGCTCGCCACGGGACACCATCACCGGCTCAACGCCATCATCAACGAG 108  
Qy 62 GCGTTCCTCGGCATTCCCTTTGCGAGCGCGCGGTGGCAACCTCGCGCTTCAAGGACCTT 121  
Db 109 GCGTTCCTCGGCATTCCCTTTGCGAGCGCGCGGTGGCAACCTCGCGCTTCAAGGACCTT 168  
Qy 122 GTGCGGTACTGTGGCTCGCTCAACGGCGAGAACTT-----ACTTACGGCGCGTGATG 174  
Db 169 GTGCGGTACTGTGGCTCGCTCGATGCGGACCAAGTTACGCTGTACGGCGCGCTGTGATG 228  
Qy 175 CAGCAACACCGAGGCGAGTTTGAAGAGAACCTTGGCAAGCGCACTCGACTTGGTG 234  
Db 229 CAGCAACACCGAGGCGAGCTTACAGGAGAACCTTCCCAAGGACGCGCTCGACTTGGTG 288  
Qy 235 ATGCACTCAAGGTGTTCAGGCGGTGTTCCTCCAGAGTGAAGTGCCTTCAACATCAAC 294  
Db 289 ATGCACTCAAGGTGTTCAGGCGGTGTTCCTCCAGAGTGAAGTGCCTTCAACATCAAC 348  
Qy 295 GTGGTGGCGCGCGGCGACCAAGGGGGCGCAACCTCCCGGTATGTCTCGACTTTT 354  
Db 349 GTGGTGGCGCGCGGCGACCAAGGGGGGTGCCAACCTCCCGGTATGTCTCGACTTTT 408  
Qy 355 GGCGGTGGGTTCAGATCGGCAGCCCAACCATCTTCCTCCCGCCAGATGGTCAACAG 414  
Db 409 GGCGGGGGGTTCAGGTGGGTGGCACCAAGCACTTCCTCCCGCCAGATGATCAACAG 468

Qy 415 AGTGTCTCATGGGCAAG--CCATCATCCAGTGGCGCTCAACTACCGTGTGTGCTCGTGG 473  
Db 469 AGCATTGCATGGGCAAGCCCATCATCCAGTGGAGGTCAACTACCGGTGTGTGCTGGTGG 528  
Qy 474 GGGTCTTGGCTGGTGTATGACATCAAGGCGGAGGAGCGGGGAACCGCGCTTGAAGGAC 533  
Db 529 GGGTCTTGGCTGGCGACGAGATCAAGGCGGAGGAGTCCCAACCGCGTTCGAAGGAC 588  
Qy 534 CAGCGTTTGGGCGATGAGTGGGTGGCAGACAATTGCCGGGTTCGGCGCGGACCCGAGC 593  
Db 589 CAGCGTTTGGGCGATGAGTGGGTGGCGGACAATTGCCGGGTTCGGCGCGGACCCGACC 648  
Qy 594 AAGGTGA-CATCTTTGGCGA---GGCGGCGAGCATGTCCGTGTGTGTGCAACCTCATCTGG 649  
Db 649 AAGGTGACCATCTTTGGCGAGCTGGCGGCGAGCATGTGTGCTCATGTGCGACATCTCTGG 708  
Qy 650 AAGCAGCGGCAACACAGTACAAAGGCAAGCGGTGTTCGGCGCGGCGCATCATGCA---G 706  
Db 709 AAGCAGCGGCAACACAGTACAAAGGCAAGCGGTTCGGCGCGGCGCATCATGCAAGCTG 768  
Qy 707 GGAGCCATGTGTC---GGACCCGCTGGAGCGGCACTGCGCAACGAGATCTACGACCTC 763  
Db 769 GGGGCCATGTGCGCTGGAGCGGCTGGAGCGGCTTACGGCAACGAGATCTTTGACCTC 828  
Qy 764 TTTGTCTGAGTGTGCTGTGGCGGCGAGCAAGCTCGCGTGTGTGGCAGTGC- 822  
Db 829 TTGGCGTGAACCGCGGCTGGCGAGCGGCGAGCAAGCTTGGCTGTGGCGGTGTG 888  
Qy 823 --GAGGCAACCTTGTCTGATGCGCAACAAACACTTCCTGGTTCCTGGGTACTCTCTG 880  
Db 889 CTGAGCGACACGTTGGAGGAGCGCAACAAACACCCCTTGGTGTCTTGGCGTACTCTCTG 948  
Qy 881 TTGGCGGTG-----TACTCCCGCGCGGCGAGCAAGCAACATCACCGATGACATGTAACAG 934  
Db 949 TTGGCGGTGCTGTACTCTCCCGCGCGGCGGAGCAATCACCGACGACATGTAGGCC 1008  
Qy 935 TTGGTGGCGGCGGCAAGTATGCAAGCTTCCCGTGTATCATTTGGCGACAGCAACGAG 994  
Db 1009 TTGGTGGCGGAGGCAAGTATGCAACATCCCTGTGTATCATCGCGACAGCAACGAG 1068  
Qy 995 GGCACATCTTTG-----GCTTTGAACGTGACCAAGATGCTCAGGCCCGTGTCTTAC 1047  
Db 1069 GGCACCTTCTTTGGCACCCCTGTGTGTAACGTGACCAAGTGTCCAGCGCGCGCGAGTAC 1128  
Qy 1048 TTCAAGCA--GTTTATCCAGCGCGAGCGAGATGACACTTGTATGGCGGTATC 1104  
Db 1129 TTCAAGCACTGTTTGTCCAGCGCGAGCGGAGATGACACGTGTATGAGCGGTATC 1188  
Qy 1105 CCCAGGACATCACCCAGGG---TCCGTTTCGACACGGTT-----CAAGCTCACCCCG 1154  
Db 1189 CCGGCGCATCACCCAGGCGCTGCGTTTCGACACGGGTATTCTCAACGCGCTCACCCCG 1248  
Qy 1155 CAGTTCAAGAGATC---CGGTGTCTGGGAGCTTGCATTCATCCAGCGCGCGCGTAC 1211  
Db 1249 CAGTTCAAGAGATTCCTGGCGGTGTCTGGCGACCTTGGCTTTACGCTGTCTGCTGCTAC 1308  
Qy 1212 TTCTCAACCACTTCAGGCGGCGACCAAGTACTCGTTCCTC-----AAGCAGCTCGGG 1265  
Db 1309 TTCTCAACCACTACAGCGGCGCACCAAGTACTCATTCCTCTGAAGCAGCTCTCTGGGC 1368  
Qy 1266 TTGCCAATCATGGGCACTTTCATGCCAACGACATTTGTGTGGCAGACTACTTTTGGGA 1325  
Db 1369 TTGGCGGTCTCGGAACGTTTCACTCCAAACGACATTTGTCTTCAGGACTACTTTTGGGC 1428  
Qy 1326 AGGGCAGCGGTATCTAACAAACCGGTTTATCGCGTTCGCAACCGACTTGAACCCCAAC 1385  
Db 1429 AGCGGTCTCATCTTACAAACCGGTTTCAITGCGTTTGGCAGGACTTGGACCCCAAC 1488  
Qy 1386 ACCCGGGGTGTGTGTGAACCTGGCCCAAGTACACCGAGC-----CAGGCAACAAAC 1439  
Db 1489 ACCCGGGGTGTGTGTGAAGTGGCCCGGAGTACACCGAGCGCTGAGCTGGGCAACAAAC 1548  
Qy 1440 TTGATGATCAACCGCTTGGGCTTGTATACCGGCAAGGACAACTTCGCGCACCGTGGC 1499

```

Db      1549 TTGATGATCAAGCGCTTGGGCTTGATACCGGCAAGGACAATCTCCGACCGCGC 1608
Qy      1500 TACGACGGTTGATGACCAACCG 1523
Db      1609 TACGACGGTTGTTCTTCCAACCG 1632

RESULT 7
CCLIP1      1733 bp      DNA      linear      PLN 27-APR-1993
LOCUS      C.cylindracea Lip1 gene for lipase.
DEFINITION      X64703
ACCESSION      X64703.1 GI:2544
VERSION      Lip1 gene.
KEYWORDS      Candida cylindracea
SOURCE      Candida cylindracea
ORGANISM      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
              Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE      1 (bases 1 to 1733)
AUTHORS      Longhi,S., Fusetti,F., Grandori,R., Lotti,M., Vanoni,M. and
              Alberghina,L.
TITLE      Cloning and nucleotide sequences of two lipase genes from Candida
              cylindracea
JOURNAL      Biochim. Biophys. Acta 1131 (2), 227-232 (1992)
MEDLINE      92305068
PUBMED      1610906
REFERENCE      2 (bases 1 to 1733)
AUTHORS      Longhi,S.
TITLE      Direct Submission
JOURNAL      Submitted (28-FEB-1992) S. Longhi, Universita' degli Studi di
              Milano, Via Celozia 26, 20133 Milano, ITALY
COMMENT      See also x64704
              X66006, X66007 and X66008 are related sequences in the description
              of C.cylindracea Lip1 and Lip2 sequences (x64703, x64704).

FEATURES
source
1..1733
    /organism="Candida cylindracea"
    /mol_type="genomic DNA"
    /strain="ATCC 14830"
    /db_xref="taxon:44322"
    /clone="A"
1..1650
    /gene="LIP1"
1..1650
    /gene="LIP1"
    /EC_number="3.1.1.3"
    /codon_start=1
    /product="triacylglycerol lipase"
    /protein_id="CAA45957.1"
    /db_xref="GI:1325987"
    /db_xref="GOA:P20261"
    /db_xref="SWISS-PROT:P20261"
    /translation="MELALALLIASVAAPATLAIANGDTTGLNAINAEPLGIPPA
    EPPVGNLRFKDPVPYSGSLDQKFTLYGLPCMQNPEGYEENLPKALDLVMSKVF
    EAVPLSDCLTIVVRPGTKAGANLFWLWIFGGFVGSTSTFPDPAQMTKSIAM
    KPIITHSVNVRVSWGLAGDEIKAEGSANAGLDORLGMOWADNIAAFGGDPTKV
    TIFGLAGSMVMCHILWANDGNTYKGPLFRAGIMOLGAMVPLDAVDGIYGNIEFL
    LASNAGCSADKLACRGVLSDTLEDATNTPGFLAYSSRLILYLPRPDGWNIDDM
    YALVREGKIANIPIVIGQNDDEGTFPGILLNVITDAQAREIFFQLFVRASDAIEDIL
    MTAYPGDITQGLPDTIGTINALTQFGRILAVLIGDLFTLARRVPLNHYTGTKYSFL
    LKQILGLPVLGTFFHSNDIVFDYLLIGSGLIYNNAFIAPATDLDPNTAGLLRWKPEYT
    SSQILGNLMMINALGLYTGKDNFTAGYDALFSNPPLFFV"
1..45
    /gene="LIP1"
    /note="triacylglycerol lipase"
46..1647
    /gene="LIP1"
    /product="triacylglycerol lipase"
    /EC_number="3.1.1.3"

sig_peptide
mat_peptide

ORIGIN
Query Match      66.7%; Score 1022; DB 8; Length 1733;

```

```

Best Local Similarity      84.7%; Pred. No. 7.4e-159;
Matches 1342; Conservative      0; Mismatches 180; Indels      62; Gaps      15;

Qy      2      CCCACCCGCAAGCTCGCCCAAGCGGACACCATCACCGGTCTCAAGGCCATCATCAACGAG 61
Db      49      CCCACCCGCAAGCTCGCCCAAGCGGACACCATCACCGGTCTCAAGGCCATCATCAACGAG 108
Qy      62      GCGTTCCTCGGCATTTCCTTTGCCGAGCCGCGGTGGGCAACCTCCGCTTCAAGAGACCT 121
Db      109     GCGTTCCTCGGCATTTCCTTTGCCGAGCCGCGGTGGGCAACCTCCGCTTCAAGAGACCT 168
Qy      122     GTGCGGTACTCTGTGCTCGCTCAACCGGCCAGAGTT-----ACTTACGGCCCGGTGATG 174
Db      169     GTGCGGTACTCTCGGCTCGCTCGATGGCCAGAGTTTCACGCTGTACGGCCCGCTGTGCTG 228
Qy      175     CAGCAGAACCCCGAGGGCAGCTTTTGAGAGAACCTTGGCAGAGCGGCATCGACTTGTGTG 234
Db      229     CAGCAGAACCCCGAGGGCAGCTTACGAGAGAACCTTCCCAAGGAGCGCTCGACTTGTGTG 288
Qy      235     ATGCAGTCCCAAGGTGTTCCAGGCGGTGTTCCCGCAGAGTGAGGACTGCCTCACCATCAAC 294
Db      289     ATGCAGTCCCAAGGTGTTTGAGGCGGTGCTGCCGCTGAGCAGGAGTGTCTCACCATCAAC 348
Qy      295     GTGCTGCGGCCCGCGGACCAAGCGGCGCGCAACCTCCCGGTCACTGCTGTGATCTTT 354
Db      349     GTGCTGCGGCCCGCGGACCAAGCGGCGGTGCCAACCTCCCGGTGATGCTCTGGATCTTT 408
Qy      355     GGCGGTGGTGTGAGATCGGCAGCCCAACCATCTTCCCTCCCGCCAGATGGTCAACAG 414
Db      409     GGCGGCGGTTTGAGGTGGGTGGCACAGCACTTCCCTCCCGCCAGATGATCAACAG 468
Qy      415     AGTGTGCTCATGGGCAAG-CCATCATCAAGTGGCGGTCAACTACCGTGTGCTCGTGTG 473
Db      469     AGCATTTGCATGGGCAAGSCCATCATCCAGCTGAGCGTCAACTACCGGTGTGCTGTG 528
Qy      474     GGGTCTTGGCTGGTGATGATCAAGSCCGAGGCGGAGCGGCGGTGAGAGAC 533
Db      529     GGGTCTTGGCTGGCGACGAGTCAAGSCCGAGGCGAGTGCCAAAGCGCGGTGAGAGAC 588
Qy      534     CAGCGTTTGGCATGCACTGGGTGGCAGACAACATTTGCCGGTTCGGCGGCGACCCGAGC 593
Db      589     CAGCGTTTGGCATGCACTGGGTGGCGGAGACAATTTGGCGGTGTTGGCGGCGACCCGACC 648
Qy      594     AAGGTGA-CATCTTTGGCGA---GGCGGGCAGCATGTCCGTGTTGTGTGCCACTATCTGG 649
Db      649     AAGGTGACCATCTTTGGCGAGCTGGCGGGCAGCATGTCCGTTCATGTGCACATTTCTCTGG 708
Qy      650     AACGACGCGGACAACACGTACAAGGCGAAGCGGTGTTCCGGCGGCGCATATGCA---G 706
Db      709     AACGACGCGGACAACACGTACAAGGCGAAGCGGTCTTCCCGCGGCGCATCATGCGGTG 768
Qy      707     GGAGCCATGTTGCC---GGACCCCGTGGACGCGCAGTACGGCAACGAGATCTTACGACCTC 763
Db      769     GGGGCCATGTTGCCCTGGAGCGCGTGGACGCGCATACGGCAAGAGATCTTTGACCTC 828
Qy      764     TTTGTCTCAGTGTGCTGTGGCAGCGCCACGAGCAAGTCCGCGTGTGTCGCGAGTGC- 822
Db      829     TTGGCGTCAACCGCGGGCTGCGGAGCGCCAGCGCAAGTTCGCTGTGCTTGGCGGTGTG 888
Qy      823     --GAGCGACACTTGTCTGATGGCCACCAACACTCCCTCGGTTCCTGCGGTACTCTCCG 880
Db      889     CTGAGCGACACTTGTGGAGGACGCCCAACCAACCCCTGGGTCTTGGCGGTACTCTCCG 948
Qy      881     TTGGCGTTG-----TACTCCCGGCGCCGACGGCAAGAACATCACCGATGACATGTACAG 934
Db      949     TTGGGTGTGTTACTCTCCCGCGCGCGCGGTGAACATCACCGACGACATGTATCGCC 1008
Qy      935     TTGGTGGCGGACGGCAAGTATGCAAGCGTTCCGCTGATCATTTGGCGACAGAACGAG 994
Db      1009     TTGGTGGCGGAGGCAAGTATGCCAACATCCCTGTGATCATCGGCGACAGAACGAG 1068
Qy      995     GGCACCATCTTTG-----GCTCTTGAACGTGACCAAGATGTCTCAGGCCCGTGTCTAC 1047

```

```
Db 1069 GGCACCTTCTTTGGCACCCCTGCTGTTGAACGTGACCAACGATGCCAGGCCCGCGAGTAC 1128
Qy 1048 TTCAAGCA---GTTCAATCCAGCCAGCGACGCGAGAGTGCACACTTGTATGGCGCGTAC 1104
Db 1129 TTCAAGCAGCTGTTTGTCCAGCCAGCGACGCGAGATGCACAGTTGATGACGGCGTAC 1188
Qy 1105 CCCAGAGACATCAACCCAGGG---TCCGTTGACACAGGTT-----CAAGCTCACCCCG 1154
Db 1189 CCGGCGACATCAACCCAGGGCTGCCGTTGCACACGGGTATTCTCAACGGCCCTCACCCCG 1248
Qy 1155 CAGTTCAAGAGATC---GGGTTGCTCGGACGCTTGCATTCATCCAGCCCGCGGTAC 1211
Db 1249 CAGTTCAAGAGATCCTCGCGGTGCTCGGCGACCTTGGCTTTACGCTTGTCTCGCGTAC 1308
Qy 1212 TTCTCAACACACTTCCAGGGCGGCACCAAGTACTCGTTTCCTC-----AAGCACTCGGG 1265
Db 1309 TTCTCAACACACTACACCGGGGACCAAGTACTCATCTTCTCTGAAGAGCTCCTGGGC 1368
Qy 1266 TTGCCAATCATGGGCACTTCCATGCGCAACGACATTTGTGTGGCAGGACTACTTGTGGGA 1325
Db 1369 TTGCCGCTGCTCGGAACGTTTCCACTCCAAACGACATTTGCTTCCAGGACTACTTGTGGC 1428
Qy 1326 AGCGGAGCGTCTACTACAAACGCGTTTATCGCTTGGCACCAGCTTGGACCCCAAC 1385
Db 1429 AGCGGCTCGCTCATCTACAAACGCGTTTCAATGCGTTTGGCAGGACTTGGACCCCAAC 1488
Qy 1386 ACCGCGGGGTTGTGTGTAACCTGGCCCAAGTACACCAAGCAGC-----CAGGGCAACAC 1439
Db 1489 ACCGCGGGTGTGTGTAAGTGGCCGAGTACACCAAGCAGCTGCGAGCTGGGCAACAC 1548
Qy 1440 TTGATGATGATCAACGCTTGGGTTGTATACCGGCAAGGACAACTTCCGACCGCTGGC 1499
Db 1549 TTGATGATGATCAACGCTTGGGCTTGTATACCGGCAAGGACAACTTCCGACCGCGGC 1608
Qy 1500 TAGCAGCGTTGATGACCAACCCG 1523
Db 1609 TAGCAGCGTTGTTCTCCAACCCG 1632

RESULT 8
AX709927
LOCUS AX709927 1511 bp DNA linear PAT 10-APR-2003
DEFINITION Sequence 9 from Patent EP1288294.
ACCESSION AX709927
VERSION AX709927.1 GI:29786305
KEYWORDS
SOURCE Candida rugosa
ORGANISM Candida rugosa
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
1
Tang, S.J., Lee, G.C. and Shaw, J.F.
Recombinant candida rugosa lipases
Patent: EP 1288294-A 9 05-MAR-2003;
Academia Sinica (TW)
FEATURES
source
1..1511
Location/Qualifiers
/organism="Candida rugosa"
/mol_type="unassigned DNA"
/db_xref="taxon:5481"

ORIGIN
Query Match 65.8%; Score 1007.6; DB 6; Length 1511;
Best Local Similarity 85.7%; Pred. No. 1.8e-156;
Matches 1313; Conservative 0; Mismatches 184; Indels 35; Gaps 16;

Qy 3 CCACCGCAAGCTGGCAACGGGACACCATCATCCGCTCTCAACGCCATCATCAACGAGG 62
Db 2 CCACCGCAAGCTGGCAACGGGACACCATCATCCGCTCTCAACGCCATCATCAACGAGG 61
Qy 63 CGTTCTCGCATTCCTTTGCCAGCGCGGTTGGGACCACTCCGTTCAAGGACCCCTG 122
Db 62 CGTTCTCGCATTCCTTTGCCAGCGCGGTTGGGACCACTCCGTTCAAGGACCCCG 121
```

```
Qy 123 TGCCTACTCTGGCTCGCTCAACGGCCAGAGTTTACTTACGGCCCGTGCATGCAGCAGAA 182
Db 122 TGCCTACTCTGGCTCGCTCGATGGCCAGAGTTTACTACGGCCCGTGCATGCAGCAGAA 181
Qy 183 CCCGAGGGCACGTTTGAAGAGAACTTTGGCAAGACGGCACTCGACTTGGTGTATGCAGTC 242
Db 182 CCCGAGGGCACCTACGAGGAGAACTTCCCAAGGACGCTCGACTTGGTGTATGCAGTC 241
Qy 243 CAAGTGTTCAGGCGGTGCTTCCCAAGATGAGGACTGCTTACCATCAACGTGGTGG 302
Db 242 CAAGTGTTCAGGCGGTG-----CCGAGCGAGGACTGTCTCAACATCAACGTGGTGG 295
Qy 303 GCGCCGCGGACCAAGGCGGCGCAACCTCCGCTCATCTCTGATCTTTGGCGGTGG 362
Db 296 GCGCCGCGGACCAAGGCGGGTGCAACCTCCCGTATGCTCTGATCTTTGGCGGTGG 355
Qy 363 GTTTGAGATCGGAGCGCCCAACCATCTTCCCTCCCGCCAGATGCTCACCAAGAGTGTGCT 422
Db 356 GTTTGAGTGGTGGCACCAAGCACTTCCCTCCCGCCAGATGATCACCAAGAGCATGTC 415
Qy 423 CATGGGCAAG-CCATCATCTCAGCTGGCGCTCAACTACCGTGTTCCTCTGTTGGGTTCTT 481
Db 416 CATGGGCAAGCCCATCATCCAGCTGAGCGTCAACTACCGCTGTCTGCTGGGTTCTT 475
Qy 482 GGCTGTGTGATGACATCAAGGCCGAGGCGAGCGGGAACCGCGCTTGAAGACCAAGCGTTT 541
Db 476 GGCTGGCGCAGATCAAGGCCGAGGCGAGTGCACACCGCGTTTGAAGACCAAGCG-CT 534
Qy 542 GGGCATGCACTGGGTGGCAGACAACTATGCCGGGTTTCGGCGGACCCCGAGCAAGGTGA- 600
Db 535 GGGCATGCACTGGGTGGCGGACAACTATGGCGGTTTTCGGCGGACCCCGAAGGTGAC 594
Qy 601 CATCTTTGGCGAGCGGCGAGCATGCTCGTGTGTGCCACTCATCTGGAACGAGCGGA 660
Db 595 CATCTTTGGCGAGCGGCGAGCATGCTCGTGTGTGCCACTCATCTGGAACGAGCGGA 654
Qy 661 CAACAGCTACAAGGCAAGCGCTTGTTCGCGGGGCGATCATCGAGGAGCCATGTTGCC 720
Db 655 CAACAGCTACAAGGCAAGCGCTTTCGCGGGGCGATCATCGAGGAGCCATGTTGCC 714
Qy 721 GGAACCGGTGGAGCGGACGCTACCGCAACAGAGATCTACGACCTTCTTCGAGTGTGG 780
Db 715 GGAACG-GTGGACGGCTC-TACGGCAACAGAGATCTTTCGCTTCG-CCGCGG 771
Qy 781 CTGTGGCAGCGGCGAGCGACAGCTCCGCTGCTTGGCAGTGCAGGACACCTTGTCTGA 840
Db 772 CTGCGGAGCGGCGAGCGACAGCTTTCGCTGCTTGGCGGTTGTGAGCGACACGTTGGAGGA 831
Qy 841 TGCCACCAACACACTCTCTGGGTTCCTGGCTACTCTCTGTTCCGCTTGTACT-CCGCGC 899
Db 832 CGCCACCAACACACCTCTGGGTTCCTGGCTACTCTCTGTTCCGCTTGTACTCTCCCGGC 891
Qy 900 CGACCGGCAAGACATCAACCGATGACATGTACAAGTTGGTGGCGGACGCAAGTATGCA 959
Db 892 CGACCGGCTGAACATCAACCGACATGTGC---CTTGGTTCGCGAGGCAAGTATGCA- 947
Qy 960 GCGTTCCGCTGATCATTTGGGACCAAGACGAGGCGACCATCTTTGGC-TCTTGAACG 1018
Db 948 ---CTCCTGTGATCATTCGGCGACCAAGACGAGGCGACCTTCTTTGGCACCTTGAACG 1004
Qy 1019 TGACCAAGATGCTCAGGCCCGTGTACTTCAAGCAGTTCATCCACGCGCCAGCGCGG 1078
Db 1005 TGACCAAGATGCGAGGCC---GAGTACTTCAGCAGTTTGTCCACGCGAGCGCGG 1061
Qy 1079 AGATCGACACCTTGTATGGGCGGTACCCCGAGGACATCACCCAGGTCGCTTCGACACG 1138
Db 1062 AG-TCGACAGTTGATGACGGGTACCCCGACATCACCCAGGCGCTTCGACACGCGTT 1120
Qy 1139 TTCAAG--CTACCCCGCAGTTCAAGAGATCGCGGTGCTCGGCGACCTTGTGATTCATC 1196
Db 1121 CTCAAGCCCTCAACCCCGCAGTTCAAGAGATCGCGGTGCTCGGCGACCTTGTGTTACG-- 1178
```



QY 1197 CAGCGCGCGCTACTTCTCAACCACTTCCAGGCGCGCAACCAAGTACTCTGTTCTCTCAAG 1256  
 Db 1179 -----CGTCTACTTCTCTCAACCACTTCCAGGCGCGCAACCAAGTACTCTCTCAAG 1233  
 QY 1257 CAGCTCGGTTGCCAATCATGGGCACTTCCATGCCAAGCAATGTTGTGGCAGGACTAC 1316  
 Db 1234 CAGCTCGGTTGCCGTTCTCGAAGCTTCCACTCCAACGACATTTGTCTTCAGGACTAC 1293  
 QY 1317 TTGTTGGAGCGGAGGCTCATCTACAAACCGGTTTATCGGTTTCGCCACCGACTTG 1376  
 Db 1294 TTGTTGGGACGGCTCGCTCATCTACAAACCGGTTTATCGGTTTCGCCACCGACTTG 1353  
 QY 1377 GACCCCAACACCGGGGTTGTTGTGAATCGGCCCAAGTACACAGCAGCAGGCGCAAC 1436  
 Db 1354 GACCCCAACACCGGGGTTGTTGTGAATCGGCCCAAGTACACAGCAGCAGGCGCAAC 1413  
 QY 1437 AACTTGATGATCAACCGCTTGGGCTTGTATACCGGCAAGGCAAACTTCGGACCGCT 1496  
 Db 1414 AACTTGATGATCAACCGCTTGGGCTTGTATACCGGCAAGGCAAACTTCGGACCGCT 1472  
 QY 1497 GGCTACGACGGTTGATGACCAACCGGTTCTT 1528  
 Db 1473 GGCTACGACGGTTGTTCTTCAACCGCGGTT 1504

RESULT 9

LOCUS A81171 1647 bp DNA linear PAT 21-JAN-2000  
 DEFINITION Sequence 1 from Patent WO9914338.  
 ACCESSION A81171  
 VERSION A81171.1 GI:6731630  
 KEYWORDS  
 SOURCE unidentified  
 ORGANISM unidentified  
 REFERENCE 1 (bases 1 to 1647)  
 AUTHORS Schmidt-Dannert, C. and Schmid, R.  
 TITLE TOTAL SYNTHESIS AND FUNCTIONAL OVEREXPRESSION OF A CANDIDA RUGOSA  
 LIP1 GENE CODING FOR A MAJOR INDUSTRIAL LIPASE  
 JOURNAL Patent: WO 9914338-A 1 25-MAR-1999;  
 SCHMIDT DANNERT CLAUDIA (DE); SCHMID ROLF (DE)  
 FEATURES  
 source  
 1..1647  
 /organism="unidentified"  
 /mol\_type="unassigned DNA"  
 /isolate="985229"  
 /db\_xref="taxon:32644"

ORIGIN

Query Match 65.4%; Score 1002.4; DB 6; Length 1647;  
 Best Local Similarity 84.3%; Pred. No. 1.3e-155;  
 Matches 1336; Conservative 0; Mismatches 186; Indels 62; Gaps 16;  
 QY 2 CCACCGCAAGCTCGCCAAACGGGACACCATCAACGGTCTCAACGCCATCATCAACGAG 61  
 Db 49 CCACCGCAAGCTCGCCAAACGGGACACCATCAACGGTCTCAACGCCATCATCAACGAG 108  
 QY 62 GGTTCCTCGGATTCCTTTTCGGGACCGCGGTGGGCAACTCCGCTTCAAGGACCT 121  
 Db 109 GGTTCCTCGGATTCCTTTTCGGGACCGCGGTGGGCAACTCCGCTTCAAGGACCT 168  
 QY 122 GTGCGGCTACTCTGGCTCGCTCAACGGCCAGAGTT-----ACTTACGGCCCG---TGCA 174  
 Db 169 GTGCGGCTACTCTGGCTCGCTCGATGCGCAGAGTTCAACGAGTCTACGGCCGAGCTGATG 228  
 QY 175 CAGCAGAAACCCGAGGCGAGTTTGAAGAGAACCTTGGCAGAGCGGCACTCGACTTGGTG 234  
 Db 229 CAGCAGAAACCCGAGGCGAGTTTGAAGAGAACCTTGGCAGAGCGGCACTCGACTTGGTG 288  
 QY 235 ATGCACTCCAAAGGTTCAGGCGGTGCTTCCCGAGGTGAGGACTGCTTCAACATCAAC 294  
 Db 289 ATGCACTCCAAAGGTTCAGGCGGTGAGGCGGTGAGGCGGTGAGGACTGCTTCAACATCAAC 348

QY 295 GTGTCGCGCGCGCGCGGCAACAGGCGGCGCGCAACCTCCCGGTCACTGCTCTGGATCTTT 354  
 Db 349 GTGTCGCGCGCGCGCGGCAACAGGCGGCGCGCAACCTCCCGGTGATGCTCTGGATCTTT 408  
 QY 355 GCGGTCGGTTGAGATCGGCGAGCCCGACCATCTTCCCTCCCGCCAGATGGTCAACCAAG 414  
 Db 409 GCGGTCGGTTGAGGTTGGGTTGGGCGAGCCCTTCCCTCCCGCCAGATGATCAACCAAG 468  
 QY 415 AGTGTCTCATGGGCAAG-CCATCATCCAGTGGCGGTCAACTACCGTGTTCCTCTGTTGG 473  
 Db 469 AGCATTTGGGATGGGCAAGCCCATCATCCAGTGGCGGTCACTACCGGCTGCTGCTGG 528  
 QY 474 GGGTTCTTGGTGTGTGATGATCAAGGCGGAGGCGAGCGGAGACCGCGCTTGAAGAC 533  
 Db 529 GGGTTCTTGGTGTGTGATGATCAAGGCGGAGGCGAGTGCACACCGCGGTTGAAGAC 588  
 QY 534 CAGGTTTGGCATGAGTGGTGGCAGACAAATTCGCGGTTTCGGGCGGCGACCGGAGC 593  
 Db 589 CAGGCTTGGGATGAGTGGTGGGCGAGCAAACTTGGCGGTTCGGGCGGCGACCGGAGC 648  
 QY 594 AAGGTGAC-ATCTTTTGGCGAG---GCGGGCAGCATGTCCGTTGTGTGCACTCATCTGG 649  
 Db 649 AAGGTGACATCTTTTGGCGAGCGCGGCGAGCATGTCCGTCATGTGCGCATCTCTGG 708  
 QY 650 AACGACGCGGCAACACAGTACAGGCGAGCGGTTGTTCCGGCGGCGCATCATGCGAGC 706  
 Db 709 AACGACGCGGCAACACAGTACAGGCGAGCGGCTCTTCCGGCGGCGCATCATGCGAGC 768  
 QY 707 GGAGCCTATGGTGGCG---GACCCGTTGAGCGGCGAGTACGCGCAACAGCATCTACGACCTC 763  
 Db 769 GGGCGATGGTGGCGGCGAGCGCGTGGACCGGCTATACGCGCAACGAGATTTTGGACCTC 828  
 QY 764 TTTGTCTCGAGTGTGCTGTGGCAGCGCCAGCAAGCTCGGTTGTGGCGAGTGGC 823  
 Db 829 TTGGGCTGAACGCGGCGGTGGCGAGCGCCAGCAAGCTCGCGTGTGGCGGCTGTG 888  
 QY 824 ---AGCGACACTTGTCTGATGCGCAGCAACCAACTCTCTGGTCTTGGCGTACTCTCG 880  
 Db 889 AGCAGCGACACTTGTGGAGCGCGCCACCAACACCCCTGGGTTCTTGGCGTACTCTCG 948  
 QY 881 TTGGGTTGTACT-----CCCGCCCGACGCGCAAGACATCACCGATGACATGATCAAG 934  
 Db 949 TTGGGTTGAGTACTCTCCCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1008  
 QY 935 TTGGTTCGCGAGCGCAAGTATGCAAGCGTTCCGTTGATCATTTGGCGACCGAGCAACGAG 994  
 Db 1009 TTGGTTCGCGAGCGCAAGTATGCAAGCATCCCTGTGATCATCGGCGACCGAGCAACGAG 1068  
 QY 995 GGCACCATCTTTGGC-----TCTTGAACGTGACCAAGATGCTCAGGCGCGTCTTAC 1047  
 Db 1069 GGCACCTTCTTTTGGCACCGAGCTTGAACGTTGACCGGATGCGCGGCGCGCGGAGTAT 1128  
 QY 1048 TTCAAGCAG---TTTCATCCAGCGCGAGCGCGGAGATCGACACCTTGTATGGCGGTAC 1104  
 Db 1129 TTCAAGCAGAGTTTGTTCACCGCGAGCGCGGAGATCGACACGTTGATGACGCGGTAC 1188  
 QY 1105 CCCCAGGACATCAACCCAGG---GTCCGTTGCAACCGGTT-----CAACGCTCACCOCG 1154  
 Db 1189 CCGCGGACATCAACCCAGGCGCGGTTTCGACACCGGTTATCTCAACGCCCTCAACCCG 1248  
 QY 1155 CAGTTCAAGAGAT---CGCGGTGTTCGCGGACCTTGCATTCATCCAGCGCGCGCTAC 1211  
 Db 1249 CAGTTCAAGAGATCAGCGCGGTGTTCGCGGACCTTGGGTTTACGTTGCTGTCGCTAC 1308  
 QY 1212 TTCTCTCAACCACTTCCAGGCGCGCACCAAGTACTGTTCTCTCA-----AGCAGTCTGG 1265  
 Db 1309 TTCTCTCAACCACTTACACCGCGCGCACCAAGTACTGTTCTTCTCAGCAAGCAGCTCAGCG 1368  
 QY 1266 TTGCCAATCATGGGCACTTCCATGCGCAACGAGATTTGTGTGGAGGACTACTTGTGGGA 1325  
 Db 1369 TTGGCGGTGCTCGAAGCTTCCACTCCAACGAGATTTGCTTCCAGGACTACTTGTGGGC 1428  
 QY 1326 AGCGGCGAGCTCATCTTACAAACGCGTTTATCGCGTTTCGCCACCGACTTGGACCCCAAC 1385



Db 1429 AGCGGCTCGCTCATCTACAAACGCGTTCAITTCGCTTTGCCACGACTTGGACCCCAAC 1488  
QY 1386 ACCGGGGGTTGTTGGTGAACAGGCGCAAGTACACACGAGC-----CAGGSCAACAC 1439  
Db 1489 ACCGGGGGTTGTTGGTGAAGTGGCCGAGTACACGAGCAGCAGCAGGCGCAACAC 1548  
QY 1440 TTGATGATGATCAACGCGCTTGGCTTTGTATACCGGCAAGGACAACTTCCGACCGCTGGC 1499  
Db 1549 TTGATGATGATCAACGCGCTTGGCTTTGTATACCGGCAAGGACAACTTCCGACCGCGCGC 1608  
QY 1500 TACGAGCGGTTGATGACCAACCG 1523  
Db 1609 TACGAGCGGTTGTTCTCCAACCG 1632

RESULT 10  
AX709923  
LOCUS AX709923  
DEFINITION Sequence 5 from Patent EP1288294.  
ACCESSION AX709923  
VERSION AX709923.1 GI:29786303  
KEYWORDS  
SOURCE Candida rugosa  
ORGANISM Candida rugosa  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE 1  
AUTHORS Tang, S.J., Lee, G.C. and Shaw, J.F.  
TITLE Recombinant candida rugosa lipases  
JOURNAL Patent: EP 1288294-A 5 05-MAR-2003;  
Academia Sinica (TW)  
FEATURES  
1..1548  
Location/Qualifiers  
/organism="Candida rugosa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:5491"

ORIGIN  
Query Match 65.2%; Score 998.4; DB 6; Length 1548;  
Best Local Similarity 83.8%; Pred. No. 5.9e-155;  
Matches 1301; Conservative 0; Mismatches 221; Indels 30; Gaps 14;

QY 1 GCCACCGCAAGCTCGCAACGGGACACCATCACCGGTCTCAACGCCATCATCAACGA 60  
Db 1 GCCACTGCCACGCTCGCAACGGGACACCATCACCGGTCTCAACGCCATCATCAACGA 60  
QY 61 GCGGTTCTCGGATTCCTTTGCGAGCGCGGGTGGCAACCTCGCTTCAAGAACCC 120  
Db 61 GCGGTTCTCGGATTCCTTTGCTCAGCGCGGGTGGCAACCTCGCTTCAAGCGCGC 120  
QY 121 TGTGCGTACTCTGGCTCGCTCAACGGCGAGAAGTT--ACTTACGGCGCGTGCAATG 177  
Db 121 TGTGCGTACTCGGCTCTCTAATGGTCAGAAGTTTACTGTATGGCTTGATGCAATG 180  
QY 178 CAGAACCCCGAGGCGCGTTTGAAGAGAACCTTGGCAAGCGCACTCGACTTGGTGATG 237  
Db 181 ATGAACCCATTTGGCAACTGGGACTCCTCGCTTCCCAAGGCTGCCATCAACT--TGATG 237  
QY 238 CAGTCAAGGTGTTCCAGCGGTGCTTCCCGAGAGTGAAGTGCCTCCACATCAACGTTG 297  
Db 238 CAGTCCAACTCTTCCAGCGGTGCTTCTTAAGCGGAGGAGTGTCTCAACATCAACGTTG 297  
QY 298 GTGCGCGCGCGGACCAAGCGGCGGCAACCTCCCGTATGCTCTGGATCTTTGGC 357  
Db 298 GTGCG--GCCGGGACCAAGCGGCGGTCACCTCCCGTATGTTGGATTTTGGC 354  
QY 358 GTGGGTTTGAATCGGCGCGCCACCATCTTCCCTCCCGCGCAGATGGTCAACAGAGT 417  
Db 355 GCGGGGTTGAGTTGGCGGCTCCAGTCTCTTCCCTCCCGCAGATGATCAACCGCCAGC 414  
QY 418 GTGCTCATGGCAAG--CCTATCATCAAGTGGCGGTCAACTACCGGTTGCCCTGTTGGGG 476  
|||||

Db 415 GTGCTTATGGGCAAGCCCATCATCCAGCTGAGCATGAATACCGCGTTGCTTCGTGGGG 474  
QY 477 TTCTTTGGCTGGTGAATACATCAAGCCGAGGCGAGCGGAAAGCCGGCTTTGAAGAGAC 536  
Db 475 TTCTTTGGCTGGTTCAGACATCAAGCCGAGGCGAGCGGAAAGCCGGCTTTGACAGCA 534  
QY 537 CGTTTGGGCATCGAGTGGGTGGCAGAAACATTTGCGGGTTCGGGCGGCAACCGAGCAAG 596  
Db 535 CGCTTGGGTTTGCAGTGGGTGGGACAAATTTGCGGGTTCGGGCGGACCGTCCAAG 594  
QY 597 GTGA--CATCTTTGGCGA--GGCGGGAGCATGTCGTTGTTGTCGCCACTCATCTGGAAC 654  
Db 595 GTGACCATCTTTGGTGAAGGCGGCGAGCATGCGGTAATGTGTACGCTCTCTGGAAC 654  
QY 655 CGGCGACAAACGTTACAAAGGCAAGCCGTTGTTTTCGCGCGGCGCATCATCGAGGAGCCAT 714  
Db 655 CGGCGACAAACGTTACAAAGGCAAGCCGTTGTTTTCGCGCGGCGCATCATCGAGGAGCCAT 714  
QY 715 GGTGCC--GGACCCGTTGGACGGCAAGTACGGCAACGAGATCTACGACCTCTTTGTCTGA 773  
Db 715 GGTGCCGGGACCCGTTGGATGGGCCCTACGGCAACGAGATCTACGACAGGTGTTGCTT 774  
QY 774 GTGCTGGCTGTGGCAGCGCCAGGCAAGCTCGCGTCTTGGCGAGTG--CGAGCGACAC 832  
Db 775 CAGCCGCGCTGTGGCAGTGCAGCGCAAGCTCGCGTCTTGGCGAGCATCGAACACAA 834  
QY 833 TTGCTCGATGCGCACCAACAACTCTCTGGGTTCTTGGCGTACTCTCTGTTGGGTTG-- 889  
Db 835 CTCCTCCAGGCCACAGCGACACTCGGGGGCTTGGCGTACCCCTCGTTGGTGGTT 894  
QY 890 -TACTCCCGGCGCGCAGCAAGAAATCATCACCGATGACATGTAACAAGTTGGTGGCGAG 948  
Db 895 TCTCCCGGCGCGCAGCGCACCTTCATCACCGATGACATGTTCAAGTTGGTGGCGAGCG 954  
QY 949 CAAATGATCAACGCTTCCGTCATCTTGGCGACAGCAACGAGGCGACCATCTTTG-- 1007  
Db 955 CAAATGTCGCAACGTTTCCGGTGTATCATTTGGCGACAGCAACGAGGCGACAGTTTGC 1014  
QY 1008 GCTCTTGAACGTGACACCAATGCTCAGGCCCGTCTTACTTCAAGCAGTTTCATCCACGC 1067  
Db 1015 GTTGTGAACGTGACTACGATGCTCAGGACGCGCAGTACTTCAAGGATTCATCCACGC 1074  
QY 1068 CAGCGACGCGGAGATCGACACCTTGTATGGCGCGTACCCCGAGGACATCACCCAGGTC 1127  
Db 1075 CAGCGACGCGGAGATCGACACCTTGTATGGCGCGTACCCCGAGGACATCACCCAGGTC 1134  
QY 1128 GTTCCAGACCG-----TTCAACGC--TCACCCCGCAGTTCAAGAGAT--CGCGTGT 1177  
Db 1135 GTTCCAGACCGGATCTTCAACGCCATCACCCCGCAGTTTCAACCGATTCAGCGGTGT 1194  
QY 1178 CGGCGACCTTGCATTCATCCAGCGCGCGCTACTTCTTCAACCATCTTCCAGGCGGCGAC 1237  
Db 1195 TGGTGAACCTTGGTTCACTCTCCCGCGGCTACTTCTTCAACCATCTTCCAGGCGGCGAC 1254  
QY 1238 CAAATGATCTGTTCTTC--AAGCAGCTCGGTTGCCAATCATGGGCACTTCCATGCAACG 1296  
Db 1255 CAAATGATCTGTTCTTCTTCAAGCAGCTTGGTTTGGCGGTGATTGGCACCCACCGCAACG 1314  
QY 1297 ACATTTGTGGCAGACTACTTGTGGGAAGCGGCGAGCGTCACTTACAAACAGCGTTTA 1356  
Db 1315 ACATTTGTGGCAGACTTTTGTGGTGAAGCAGCGCGCTGTACAAACAGCGTTTA 1374  
QY 1357 TCGCGTTGCCACCGACTTGGACCCCAACACCGCGGGGTTTGTGTGAACCTGGCCCAAGT 1416  
Db 1375 TTGCTTTTGGCAAGCACTCGACCCGCAAGCGCGGTTTGTGTGAACCTGGCCCAAGT 1434  
QY 1417 ACACGAGCGCGAGGCGCAACAACTTGTATGATCAACGCGCTTGGGCTTTGTACCGGCA 1476  
Db 1435 ACACGAGCGCGAGGCGCAACAACTTGTGTGAGATCAACGCGCTTGGGCTTTGTACCGGCA 1494  
QY 1477 AGGACAACTTCCGACCGCTGCTAGCGGCTTGAATGACCAACCGGTTCTT 1528  
Db 1495 AGGACAACTTCCGACCGCTGCTAGCGGCTTGTGTGTTTACCAACCGGTTTTT 1546  
|||||

```

RESULT 11
CCLIPS
LOCUS
DEFINITION C.cylindracea LIP5 gene.
ACCESSION X66008.855942
VERSION X66008.1 GI:296937
KEYWORDS lip5 gene; lipase.
SOURCE Candida cylindracea
ORGANISM Candida cylindracea
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
1 (bases 1 to 1812)
Lotti,M., Grandori,R., Fusetti,F., Longhi,S., Brocca,S.,
Tramontano,A. and Alberghina,L.
Cloning and analysis of Candida cylindracea lipase sequences
Gene 124 (1), 45-55 (1993)
93178975
PUBMED 8440480
REFERENCE
2 (bases 1 to 1812)
Longhi,S.
Direct Submission
Submitted (06-MAY-1992) S. Longhi, Universita' degli Studi di
Milano, Via Celoria 26, 20133 Milano, ITALY
See X66006-8, X16712, X64703 & X64704
X66006, X66007 and X66008 are related sequences in the description
of C.cylindracea Lip1 and Lip2 sequences (X64703, X64704).
FEATURES
Source
Location/Qualifiers
1..1812
/organism="Candida cylindracea"
/mol_type="genomic DNA"
/strain="ATCC 14830"
/db_xref="taxon:44322"
/clone_lib="SACI genomic"
/gene="LIP5"
116..1765
/gene="LIP5"
116..1765
/gene="LIP5"
/EC_number="3.1.1.3"
/codon_start=1
/product="triacylglycerol lipase"
/protein_id="CAA46807.1"
/db_xref="GI:1325991"
/db_xref="GOA:P32949"
/db_xref="SWISS-PROT:P32949"
/translation="MKLALALLIASVAAPATATLANGDITGLNALINEAFGLPFA
EPPVGNLRFKDPVYRGLNSQSFAYPLCMQNPBGTYEENLPKVALDLVMOSKVF
QAVLPNSEDCLITNVVRPGTKAGANLPVLMWIFGGFEIGSPTIFFPAQMVSKSLM
GKPIHVAVNRYLASPGFLAGPDIKAEGSSNAGLKQRLGMQWADNIAFGPQDSKLV
TIFGLAGSSVLCHLLWNGDNTYKGPILFRAGIMOLGAMVPLDPVDVGTGDIYDT
LVALTGSSASNKLACIRGLLTQALLDNTDTPGFLSVTSRLVLRPLPDGADITDM
YKVRDGRKASVPIIGDNDGFLGLLLNTTTEADAEAYLRKLFPHATDADITAL
KAAVPSDVTGGLPDTGLNALTPOLKRLINAVLGDLTFTLSRRYFLNHYTGPKISEL
LKQLGLPILGTHANDIVMWHFLGSGSVYNNAFATAPADLDPNTAGLLVQWFKEST
SSLQAGDNLQIISALGLYTGKDNFTAGYNALFADPLHFFV"
116..250
/gene="LIP5"
116..250
/gene="LIP5"
251..1762
/gene="LIP5"
/product="triacylglycerol lipase"
/EC_number="3.1.1.3"

sig_peptide
mat_peptide

ORIGIN
Query Match 64.1%; Score 982.4; DB 8; Length 1812;
Best Local Similarity 83.1%; Pred. No. 2.5e-152;
Matches 1318; Conservative 0; Mismatches 206; Indels 62; Gaps 15;

Qy 2 CCCACCGCGCAAGCTCGCCAAACGGCGACACCATCACCGGTCTCAACGCCATCATCAACGAG 61
Db 164 CCCACCGCGCAGCTCGCCAAACGGCGACACCATCACCGGTCTCAACGCCATCATCAACGAG 223
Qy 62 GCGTTCTCGGCATTCCTTTTCCGAGCGCGCGGTGGGCAACCTCCCGCTTCAAGACCTT 121

```

```

Db 224 GCGTTCTCGGCATTCCTTTTGGCGAGCGCGCGGTGGGCAACCTCCCGCTTCAAGACCTT 283
Qy 122 GTCCGCTACTCTCGCTCGCTCAACGGCCAGAGTTACTTACG-----GCCCTGTCATG 174
Db 284 GTCCGCTACTCGTGGGTCTCTCAACGGGTCAATCTTACCGGTACGGTCCGCTGTGATG 343
Qy 175 CAGCAGAACCCCGAGGGCAGCTTTGAAGAGAACCTTTGGCAAGACGGCACTCGACTTGGTG 234
Db 344 CAGCAGAACCCCGAGGGCAGCTTACGAGGAGAACCTTCCCAAGGTGGCGTTGACTTGGTG 403
Qy 235 ATGCAGTCCAAAGTGTTCAGGGCGTCTTCCAGAGTAGAGTAGTGCCTCACCATCAAC 294
Db 404 ATGCAGTCCAAAGTGTTCAGGGGTGTTCCTCCCAACAGCAGGAGTGCCTCACCATCAAC 463
Qy 295 GTGGTGGCGCGCGCGGCACCAAGGGGGCGCAACCTCCCGGTCACTGCTCTGGATCTTT 354
Db 464 GTGGTGGCGCGCGCGGCACCAAGGGGGCGCAACCTCCCGGTCACTGCTCTGGATCTTT 523
Qy 355 GGGCGTGGGTTCAGATCGGAGCGCCCAACCATCTTCCCTCCCGCCAGATGGTCACCAAG 414
Db 524 GCGCGTGGGTTCAGATCGGAGCGCCCAACCATCTTCCCTCCCGCTCAGATGGTCTCCAAG 583
Qy 415 AGTGTGCTCATGGGCAAG-CATCATCCAGTGGCGCTCAACTACCGTGTGCTCTGTTG 473
Db 584 AGTGTGCTCATGGGCAAGCCCATCATCCAGTGGCGCTCAACTACCGTGTGCTCTGTTT 643
Qy 474 GGGTTCCTGCTGTGATGACATCAAGCGCGAGGGCGGCAACCGCGGTTCGAAGGAC 533
Db 644 GGTTCCTGCGCGTCCGGACATCAAGCGCGAGGGCGAGTCCATGCGGGCTCAGAGAC 703
Qy 534 CAGCGTTTGGGCATGCAAGTGGTGGCAGACAAATTCGCGGGTTCGCGGGCGACCCGAGC 593
Db 704 CAGCGCTTGGGCATGCAAGTGGTGGCAGACAAATTCGCGGGTTCGCGGGCGACCCGAGC 763
Qy 594 AAGTGA-CATCTTTGGCGA---GGCGGGCAGCATGTCGTTGTGTCACCTCATCTGG 649
Db 764 AAGGTGACCATCTTTTGGTGAGCTGGCGGGCAGCATGTCCGTGTGTGCCACTTCTCTGG 823
Qy 650 AACGACGGCAGACACGTACAGGGCAAGCGTGTTCGCGCGGGCATCATGCA---G 706
Db 824 AATGGCGGCGACAAACAGCTACAGGGCAAGCGGTGTTCGCGGGCGGCATCATGCGACTG 883
Qy 707 GGAGCCATGTGTC---GGACCCCGTGGAGCGGCAAGTACGSCAACAGAGATCTACGACTC 763
Db 884 GGAGCCATGTGTCGCGTGGAGCGGCAAGCTATGGAACCCAAATCTATGACAG 943
Qy 764 TTTGCTCTGAGTGTGCTGTGGCAGCGCCAGCGACAGCTCGCGTGTGTCGCGAG---T 820
Db 944 TTGTTGGCTCTGACGGGCTGACAGTGCAGCAACAAAGCTTGGTGTGCTTGGGTGCTTT 1003
Qy 821 GCGAGCGACACCTTGTCTGATGCCACCAACAACTCTCTGGTTCCTTGGCGTACTCTCTCG 880
Db 1004 CTGACTCAGGATGTGCTGATGCCCAACAGACACCTCTGGTTCCTTGTGCTGACCTCG 1063
Qy 881 TTGCGGTGTGTACT-----CCCGCGCGCAGCGGCAAGAACATCAACGATGATGATCAAG 934
Db 1064 TTGCGGTGTGTACTTCCCTCCCGCGCGAGCGCGCGCAACATCAACGATGATGATCAAG 1123
Qy 935 TTGTTGGCGGACGGCAAGTATGAAAGCGTTCCTCGTATCATTTGGGACAGAAACGACGAG 994
Db 1124 TTGTTACGCGACGGCAAGTATGAAAGCGTTCCTCGTATCATTTGGCGACCAAGACGACGAG 1183
Qy 995 GGCACCATCTTTG-----GCTCTTGAACTGTGACCAAGATGTCTAGGCGCGCTGCTTAC 1047
Db 1184 GCTCTCTGTTTGTCTCTCTGCTGTGACACACCAACCGAGGCGGAGCGCGAGGCATAC 1243
Qy 1048 TTC---AAGCAGTTTATCCAGCGCAGCGAGATCGACACTTGTATGGCGCGGTAC 1104
Db 1244 CTCAGAAAGCTGTTTCATCCACGCCACCGACCGCATATCACCGATATGAAGCGCGGTAC 1303
Qy 1105 CCCAGAGCATCCCGAGGT---CCGTTGACACGG-----TTCAACGCTCACCCCG 1154

```

```

Db 1304 CCCAGCATGTCACCCAGGCTGTCGGTTCGACACCGGCATTCTCAACGCCCTTACACCC 1363
QY 1155 CAGTTCAGAGAAATC---GGGTGCTCGGCGACTTGCATTTCATCCACGCCGCCGCTAC 1211
Db 1364 CAGTCAAGCGGATCAATGCTGCTTGGGACTCACCTTTACCTCTCGCCCGCTAC 1423
QY 1212 TTCTCAACACATTCACAGGCGGCACCAAGTACTCGTTCTC-----AAGCAGCTCGG 1265
Db 1424 TTCTCAACACATACACCGGTGGTCCCAAGTACTCGTTCTCCTCTGAAGCAGCTTCTGGG 1483
QY 1266 TTGCGATCATGGCGACTTCATGCCAAGCAGATTGTGTGGGAGACTACTTGTGGGA 1325
Db 1484 TTGCCATCTTCGGCAGCTTCCACGGAAGCAGATTGTGTGGGAGCACTTTTGTGGG 1543
QY 1326 AGCGGAGCGTCTATCAACACGCGTTCATCGCTTCGCCACCGACTTGGACCCCAAC 1385
Db 1544 AGCGGAGCGTCTATCAACACGCGTTCATCGCTTCGCCACCGACTTGGACCCCAAC 1603
QY 1386 ACCGCGGGTGTGTGTGAACCTGGCCCAAGTACACGAGAGCC-----AGGCAACAAC 1439
Db 1604 ACCGCGGGTGTGTGTGAACCTGGCCCAAGTACACGAGAGCC-----AGGCAACAAC 1663
QY 1440 TTGATGATGATCAACCGCTTGGCTGTGTACACCGGCAAGCAGCACTTCGCCACCGTGGC 1499
Db 1664 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1723
QY 1500 TAGCAGCGTGTGATGACCAACCGCTT 1525
Db 1724 TACAACGCTTGTGTGCGGACCGCT 1749

RESULT 12
CCLIP4
LOCUS C.cylindracea LIP4 gene. 1785 bp DNA linear PLN 30-JUN-1993
DEFINITION X66007.55939
ACCESSION X66007.1 Gi:296935
VERSION lip4 gene; lipase.
KEYWORDS Candida cylindracea
SOURCE Candida cylindracea
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1785)
AUTHORS Lotti,M., Grandori,R., Fusetti,F., Longhi,S., Brocca,S.,
Tramontano,A. and Alberghina,L.
Cloning and analysis of Candida cylindracea lipase sequences
Gene 124 (1), 45-55 (1993)
93178975
8440480
2 (bases 1 to 1785)
Direct Submission
Longhi,S.
Submitted (06-MAY-1992) S. Longhi, Universita' degli Studi di
Milano, Via Celoria 26, 20133 Milano, ITALY
See X66006-8, X16712, X64703 & X64704
X66006, X66007 and X66008 are related sequences in the description
of C.cylindracea Lip1 and Lip2 sequences (x64703, x64704).
FEATURES
source
1. 1785
/organism="Candida cylindracea"
/mol_type="genomic DNA"
/strain="ATCC 14830"
/db_xref="taxon:44322"
/clone_lib="SACI genomic"
99.1748
/gene="LIP4"
99.1748
/gene="lip4"
/EC_number="3.1.1.3"
/codon_start=1
/product="triacylglycerol lipase"
/protein_id="CAA46806.1"
/db_xref="GI:1325990"

```

```

/db_xref="GOA:P32948"
/db_xref="SWISS-PROT:P32948"
/translation="MKLALVLSIVSVAAPTATLANGDTITGNAIINEAFILGIPPA
QPVGMLRKPVPVPSYASLNGKFTLYGLPCLMOMNPLGNWSSLPKRAINLLMQSKLF
QAVLPNGEDCLTINVRPLGTGKPNLPMVMIFGGFVGGSSLPFPAQIMTASVLM
KPIIHVSNRYVASGFLAGPDIIKAGSNGAGLDHQRGLQWVADNIAGFGDPSKV
TIFELAGSVMVCLLMDNDGNTYNGKPLFRAAIIMQAMVPLDVPDVGPGYQIYDQ
VVASCGSASDKLACLISLNDKLPQATSDTPGALAYPSLRLELPRPDGTFTIDDM
FKLVRDGKCANYPVPIIGDQDNGSTVPALLLVNTTDAQARQYFKELFIHASDEIDTL
MAAPSDITQGLPFDGTGFNATPQPKRIAAVLGLAFILPRIPYHFGGFKYSFL
LKQLGLPVLGHANDIIVMDFLVSHSVAANNFAIFANDLPNKGALLVNWPKYT
SLQLGNLNLQINALGLTYTKDNFTAGYDALFTNPLFFV"
99.233
sig_peptide
/gene="LIP4"
234.1745
mat_peptide
/gene="LIP4"
/product="triacylglycerol lipase"
/EC_number="3.1.1.3"

```

## ORIGIN

```

Query Match 61.2%; Score 937.2; DB 8; Length 1785;
Best Local Similarity 81.7%; Pred. No. 7e-145;
Matches 1296; Conservative 0; Mismatches 228; Indels 62; Gaps 16;

QY 2 CCCACCGCCAAGCTCGCCAAACGCGCACACCATCACCGGTCTCAACGCCCATCATCAACGAG 61
Db 147 CCCACTGCCACGCTCGCCAAACGCGCACACCATCACCGGTCTCAACGCCCATCATCAACGAG 206
QY 62 GGTTCCTCGCATTCCTTTGCGAGCGCGCGGTGGGCAACTCTCGCTTCAAGAGACCT 121
Db 207 GCGTTCTCGTATTCCCTTTGCTCAGCGCGCGGTGGGCAACTCTCGCTTCAAGCGCCT 266
QY 122 GTGCGGTACTCTGGCTCGCTCAACGCCCAGAGTT-----ACTTACGCCCGTGCATG 174
Db 267 GTGCGGTACTCTGGCTCTCTCAATGTGAGAGTTTACTCTGTATGGCCCTCTGTGCATG 326
QY 175 CAGCAGAACCCCGAGGGCACGTTTGAAGAGAACCTTTGGCAAGACGGCACTCGACATTGGTG 234
Db 327 CAGATGAACCCATTGGGCAACTGGGACTCTCTCGCTTCCCAAGGCTGCCATCAACCTGTG 386
QY 235 ATGCGAGTCAAGGTGTTCAGCGGTGTTCCTCCAGAGTGAGGACTGCTCCTACCATCAAC 294
Db 387 ATGCGAGTCAAGCTCTTCCAGCGGTGTTCCTTAACGGGAGGACTGTCTCACCATCAAC 446
QY 295 GTGTCGCGCCCGCGGACCAAGCGGCGGCGCACTCTCGGTGATGCTCTGGATCTTT 354
Db 447 GTGTCGCGCCCGTGGGACCAAGCGGCGGTGCCAACCTCCCGGTGATGTGTGGATTTTT 506
QY 355 GCGGTGGGTGTGAGATCGGCAGCCCCCACTCTTCCCTCCCGCCCGAGATGTCACCAAG 414
Db 507 GCGCGGGGTGTGAGGTGGCGGCTCCAGTCTCTTCCCTCCCGCAGATGATCACCGCC 566
QY 415 AGTGTGCTCATGGGCAAG--CATCATTCACGTGGCGGTCAATACCGTGTGTGCTGTGG 473
Db 567 AGCGTGTCTTATGGGCAAGCCCATCATCCACGTGAGCATGAATACCGCGTGTGTGCTGG 626
QY 474 GGTTCCTTGGCTGATGATGACATCAAGCGGAGGCGAGCGGCAAGCGCGCTTGAAGAC 533
Db 627 GGTTCCTTGGCTGATGATGACATCAAGCGGAGGCGAGCGGCAAGCGCGTGTGACAGAC 686
QY 534 CAGCGTTTGGGATGCGAGTGGGTGGCAGACACATTTGCGCGGTTCGCGCGGACCGGAGC 593
Db 687 CAAACGTTGGGTTCAGTGGGTGGGAGCAACATTTGCGGGTTCGCGCGGACCGGTC 746
QY 594 AAGGTGA-CATCTTTGGCGA---GGCGGCGAGCATGTCTGGTGTGTGCACTCATGTGG 649
Db 747 AAGGTGACCATCTTTGGTGGAGTGGCGGCGAGCATGTCTGGTAAATGTGTGAGTCTCTGG 806
QY 650 AACGAGCGGCAACACCGTACAGGGCAAGCGGTGTTCGCGCGGCGGATCATGCA---G 706
Db 807 AACGAGCGGCAACACCGTACAGCGGAGCGGTGTTCGCGCGGCGGATCATGAGCTG 866
QY 707 GGAGCCATGGTGCC---GGACCGGTGGACGCGACGTACGGCAACGAGATCTACGACCTC 763

```

```
Db 867 GGGGCCATGTCGCGTGGACCGTGGATGGGCGCTACGGCAGCAGATCTACGACG 926
Qy 764 TTTGTCTCGAGTGTGCTGTGGCAGCGCGCAGCAGCAGTCTGGTGTGGCAGTGC- 822
Db 927 GTGCTTCTCAGCGCGTGTGGCAGTGTGGCAGCAGCAGTCTGGTGTGGCAGCATC 986
Qy 823 --GAGCAGACCTTGTCTGATGTCGACCAACCAACTCTGGTGTCTGGTGTCTGCTCG 880
Db 987 CTGAACAGCAAACTCTTCCAGGCGCAGCAGCAGTCTGGGCGCTTGGCGTACCCCTCG 1046
Qy 881 TTGCGGTG-----TACTCTCCGCGCGCAGCGCAGCAATCATCCGATGACATGACAAG 934
Db 1047 TTGCGGTGTGTTCTCCGCGCGCGCAGCGCAGCCTTCATCCGATGACATGTTCAAG 1106
Qy 935 TTGTCGCGCAGCGCAATGTCGAAGGTTCCCGTGATCATCTGGGACGACGACGACGAG 994
Db 1107 TTGTCGCGCAGCGCAATGTCGAAGGTTCCCGTGATCATCTGGGACGACGACGACGAG 1166
Qy 995 GGCACCATTTTG-----GCTCTTGAACGTGACCAAGTGTCTAGGCGCGTGTTC 1047
Db 1167 GGCACATGTTTGTGCTGCTGTGTAAGTGTGACGATGCTAGGACGCGCAGTAC 1226
Qy 1048 TTCAAG---CAGTTTCATCCAGCGCAGCGCAGGATTCGACACTTGTATGCGCGGTAC 1104
Db 1227 TTCAAGGACATGTTTCATCCAGCGCAGCGCAGGATTCGACACTTGTATGCGCGGTAC 1286
Qy 1105 CCCAGGACATCACCGAGGT---CGTTTCAGCAGG-----TTCACGC---TCAACCCG 1154
Db 1287 CCCAGGACATCACCGAGGTCTCGCGTTCGACACCGGATCTTTCACGCCATCACCCG 1346
Qy 1155 CAGTTTCAGAGAAAT---CGCGGTGTCGCGGACCTTGCATTCATCCAGCGCGCGGTAC 1211
Db 1347 CAGTTTCAGAGAAATGACGCGGTGCTTGTGACCTTGTGTTCACTCTCCCGCGCGGTAC 1406
Qy 1212 TTCCTCAACCACTTCCAGGCGCGCAGCAAGTACTGTTTCCTC-----AAGCAGCTCGG 1265
Db 1407 TTCCTCAACCACTTCCAGGCGCGCAGCAAGTACTGTTTCCTCCTTGAAGCAGCTTCTGGG 1466
Qy 1266 TTGCAATCATGCGGACCTTCATGCAAGCAGCAGCAGTGTGTCAGGACTACTTGTGGGA 1325
Db 1467 TTGCGCGTGTGTCGACCAACCGCAGCAGCAGCAGTGTGTCGAGGACTTTTGTGTGAGC 1526
Qy 1326 AGCGGACGTCATCAACAGCGTTTTCGCGTTCGCGCAGCTTGCACCGCCTGACCCCAAC 1385
Db 1527 CACAGCAGCGCGGTGTAACAAACGCGTTTATGCTTTCGCAAGCAGCTGACCCGAC 1586
Qy 1386 ACCGCGGCTTGTGTTGTAACCTGCGCCAGTACACGAGCAGC-----CAGGCGCAACAC 1439
Db 1587 AAGCGCGTGTGTTGTAACCTGCGCCAGTACACGAGCAGCTGAGCTGGGCAACAC 1646
Qy 1440 TTGATGATGATCAACGCTTGGGCTTGTACCGGAGGACAACTTGGCAGCAGCTGGC 1499
Db 1647 TTGTTGAGATCAACGCTTGGGCTTGTACCGGAGGAGAACTTCCGACCGCTGGC 1706
Qy 1500 TACGACGCTGATGACCAACCGCTT 1525
Db 1707 TACGACGCTGTTTACCAACCGCT 1732

RESULT 13
CCLIP2 2043 bp DNA linear PLN 27-APR-1993
LOCUS C.cylindracea LIP2 gene for lipase.
DEFINITION X64704
ACCESSION X64704.1 GI:2546
VERSION X64704.1
KEYWORDS LIP2 gene.
SOURCE Candida cylindracea
ORGANISM Candida cylindracea
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 2043)
AUTHORS Longhi,S., Fusetti,F., Grandori,R., Lotti,M., Vanoni,M. and
```

```
Alberghina,L.
Cloning and nucleotide sequences of two lipase genes from Candida
cylindracea
Biochim. Biophys. Acta 1131 (2), 227-232 (1992)
JOURNAL 92305068
MEDLINE 1610906
PUBMED 1610906
REFERENCE 2 (bases 1 to 2043)
AUTHORS Longhi,S.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-1992) S. Longhi, Universita' degli Studi di
Milano, Via Celoria 26, 20133 Milano, ITALY
COMMENT See also X64703
X66007 and X66008 are related sequences in the description
of C.cylindracea Lip1 and Lip2 sequences (x64703, x64704).
FEATURES
source
1..2043
/organism="Candida cylindracea"
/mol_type="genomic DNA"
/strain="ATCC 14830"
/db_xref="taxon:44322"
/clone="B"
126..132
161..1807
/gene="LIP2"
161..1807
/gene="LIP2"
/EC_number="3.1.1.3"
/coon_start=1
/product="triacylglycerol lipase"
/protein_id="CAA4598.1"
/db_xref="GI:1325988"
/db_xref="GOA:P32946"
/db_xref="SWISS-PROT:P32946"
/translaton="MKLLALGAVAAPATATLANGDTITGLNAIVNEKFLGIPFAE
PVVTLAFKPPVPSASINGQOFTLGPLCMOMPMGSEFDTLPKNARHLVLOSKLFG
VVLNDEDLITVIRPPTGRASAGLPLVLMWIFCGFELGSSLPFGDMVAKSVLMG
KPVTHSMNTRVSWGFLAGDIQNEGSHAGLHDQRLAMQWADNATAGFGDPSKVT
IYGLAGSMTFVHLVWNGDNTYNGKPLFRALIMQGMVPLDPVDGTGTGTEINQV
VASAGCSADSKLALQDLQTLQATSDTPGVLYPSLLPLPDPGDTFTIDDMY
ALVRDGVYHVPVIGQNDGEGTLGLLLANVTDAQARAYFKGLFHASDAEIDTLM
AATSDJITQGLPDTGIFNATTPQFKRILALGLDPLTARRYLFLNYVQGTGKYSFL
KQLLGLVLGTGTHNDLIWQDYLVGSGSVLYNNAPIAFANDLDPKAGLWTNWPYTS
SLQGNLMQINGLGLYTGKDNFEDAIYSALFSPNPLFFV"
161..202
/gene="LIP2"
/note="triacylglycerol lipase"
203..1804
/gene="LIP2"
/product="triacylglycerol lipase"
/EC_number="3.1.1.3"

sig_peptide
161..202
/gene="LIP2"
/note="triacylglycerol lipase"

mat_peptide
203..1804
/gene="LIP2"
/product="triacylglycerol lipase"
/EC_number="3.1.1.3"

ORIGIN
Query Match 56.6%; Score 867.6; DB 8; Length 2043;
Best Local Similarity 77.8%; Pred. No. 2.1e-133;
Matches 1233; Conservative 0; Mismatches 289; Indels 62; Gaps 13;
Qy 2 CCCACCGCCAAAGCTCGCCAAACGGGACACCATCCCGTCTCAACGCCATCATCAACGAG 61
Db 206 CCCACGGCCACCTCGCCAAACGGGACACCATCCCGTCTCAACGCCATTTGTCAACGAA 265
Qy 62 GCGTTCTCGGATTCCTTTGCCGAGCGCGCGGTGGGACCTCCGCTTCAAGGACCT 121
Db 266 AAGTTCTCGGATACCTGTTGCCGAGCGCGCGGTGGGACCGCTCCGCTTCAAGCGGCC 325
Qy 122 GTGCGGTACTCTGGCTCGCTCAACCGCCAGAGT-----ACTTACGGCCCGGTGATG 174
Db 326 GTGCGGTACTCTGGCTCGCTCAACCGCCAGCAGTTCACCTGTACGGCCGCGGTGTCATG 385
Qy 175 CAGCAGAACCCCGAGGCGACGTTTCAAGAGAACCTTGGCAGACGCGACTCGACTGTGTTG 234
Db 386 CAGATGAACCTTATGGGCTCGTTTGAGGACACACTTCCCAAGAAATCGCGGCAATTTGGTG 445
Qy 235 ATGCAGTCCAGGTTTCCAGCGCGGTGTTCCCGCAGAGTGAGGACTGCTCACCATCAAC 294
```

Db 446 CTCAGTCCAAAGATCTTCCAAAGTGGTCTTCCCAACGACGAGACTGTCTCACCATCAAC 505  
QY 295 GTGGTGCAGCGCGCGGACCAAGGCGGCGCAACCTCCCGTCTATGCTCTGGAATCTTT 354  
Db 506 GTGATCCGCGCGCGCGGACCAAGGCGGCGGCTCTCCCGTGTATGCTCTGGAATCTTT 565  
QY 355 GCGGTGGGTTTGAATCGGCAAGCCCAACCATCTTCCCTCCCGCCAGATGGTCAACAAG 414  
Db 566 GCGGTGGGTTTGAATCGGCAAGCCCAACCATCTTCCAGGAGACAGATGGTGGCCAAG 625  
QY 415 AGTGTGCTCATGGGCAAGCC-ATCATCCAGTGGCGGTGAATCTACCGTGTGCTCTGTTG 473  
Db 626 AGCGTGTCTCATGGGTAAACCGGTGATCCAGTGAAGCATGAATACCGCGTGGCGTCA 685  
QY 474 GGGTCTTGGCTGGTGAATGACATCAAGCGGAGGCGGGAACCGCGCTTGAAGGAC 533  
Db 686 GGGTCTTGGCGCGCCGACATCCAGAACGAAGGAGCGGGAACCGCGCTTGAATGAC 745  
QY 534 CAGCGTTTGGGATGAGTGGGTGGGAGACAAATTTGCCGGGTTCGCGCGGCGACCCGAGC 593  
Db 746 CAGCGTTTGGCATGAGTGGGTGGGAGACAAATTTGCTGGGTTCGCGCGGCGACCCGAGC 805  
QY 594 AAGGTGAC----ATCTTGGGAGGCGGCGAGCATGTCCGTTGTTGGCCACCTCATCTGG 649  
Db 806 AAGGTGACCATATACCGCGAGCTGGCGGCGAGCATGTCCGACCTTTGTGACCTTGTGTG 865  
QY 650 AACGACGCGGACAAACAGTACAGGCGCAAGCGGTGTTCGCGCGGCGCATCATGACAGGA 709  
Db 866 AACGACGCGGACAAACAGTACAGGCGCAAGCGGTGTTCGCGCGGCGCATCATGACAGGTG 925  
QY 710 GCC-----ATGTTGCCGACCGCGTGGAGCGGACGATGCGGCAACAGAGATCTACGACCTC 763  
Db 926 GGTGTCATGTTGCGGTGACCGCGGTGGAGCGGACGATGCGGACCGAGATCTACACACAG 985  
QY 764 TTGTCTCAAGTGTGCTGTGGCAGCGCGGAGGACAAAGCTCCGCTGCTTGGCGAG--T 820  
Db 986 GTGGTGGCTGTGCGGGGTGGCAGTGGCGGAGCAAGCTCCGCTGCTTGGCGGCGCTT 1045  
QY 821 GCGAGGACACCTTGTCTGATGCCACCAACACACCTCTGGTCTTGGCGTACTCTCTCG 880  
Db 1046 CTGAGGACACGTTGTACAGGCGGACAGGACACCGCGCGGTGTGGCGTACTCCCGTGG 1105  
QY 881 TTGCGGTG-----TACTCCCGCGCGGACGGAAGACATCACCGATGACATGTACAAG 934  
Db 1106 TTGCGGTGTCTGTATCTCCCGCGCGGACGCGACCTTCATCACCGACGATGATGCC 1165  
QY 935 TTGGTGGCGGACGGCAAGTATGCAAGCGTTCCGCTGATATTTGGCGACGAGAACGACGAG 994  
Db 1166 TTGGTGGCGGACGGCAAGTATGCAAGCGTTCCGCTGATCATCGCGGTGATCATCGCGAC 1225  
QY 995 GGCACCATCTTTG-----GCTCTTGAAGTGAACCAAGTCTCAGGCGCGTGTCTAC 1047  
Db 1226 GGCATCTTTGTTGGGCTCTCTGTTGAAGTGAACCAAGTCTCAGGACGCGGCTAC 1285  
QY 1048 TTCAAGCA---GTTATCAACGCGGACGAGCGGAGATGACACCTTTGATGGCGGTAC 1104  
Db 1286 TTCAAGCAGCTGTTTATCAACGCGGACGAGTGGAGATGACACGTTGATGGCGGTAC 1345  
QY 1105 CCCCAGGACATACCCAGGTT---CCGTTGGAACAGG-----TTCAACGCTCACCCCG 1154  
Db 1346 ACCAGGACATACCCAGGTTCTGCGGTTGCAACCGGATCTTCAATGCAATCACCCCG 1405  
QY 1155 CAGTTCAAGAGATC---CGGTGCTCGGACCTTTCATCCACCGCGCGCTAC 1211  
Db 1406 CAGTTCAAGAGATCCTGCGGTTGCTTGGGACCTTGGCTTACGCTTGGCGTCTGCTAC 1465  
QY 1212 TTCTCAACACCTTCCAGGCGGACCAAGTACTCTGTTCTC-----AAGCAGCTCGG 1265  
Db 1466 TTCTCAACCTACTACAGGCGGACCAAGTACTCTGTTCTCTGAGCAGCTTCTGGGG 1525  
QY 1266 TTGCCAATCATGGGACCTTCCATGCAACGACATTTGTGGCAGGACTACTTGTGGGA 1325

Db 1526 TTGCCGCTTTGGGCACTTCCACGCAAGACATCATCTGGCAGGACTACTTGGTGGC 1585  
QY 1326 AGCGGAGCGTCTATCTACAAACAGCGGTTTATCGCTTCGCCACCGACTTGGACCCCAAC 1385  
Db 1586 AGCGGAGTGTGATCTTAAACAGCGGTTTATCGCTTGGCCACGACTTGGACCCCAAC 1645  
QY 1386 ACCGCGGGTGTGTGTGAACTGGCCCAAGTACACGACAG-----CCAGGGCAACAAC 1439  
Db 1646 AAGCGGGGTTTGGAGCAACTGGCCACGTTACACCGAGCTGTGAGCTGGGCAACAAC 1705  
QY 1440 TTGATGATGATCAACGCGCTTGGCTTGTACCGGCAAGGACACTTCCGACCGCTGGC 1499  
Db 1706 TTGATGATGATCAACGCGCTTGGGTTGTACCGGCAAGGACACTTCCGCGCGGATGCG 1765  
QY 1500 TACGACGCGTTGATGACCAACCCG 1523  
Db 1766 TACGCGCTCTTTTCCAAACCG 1789

RESULT 14  
AX709919  
LOCUS AX709919 1469 bp DNA linear PAT 10-APR-2003  
DEFINITION Sequence 1 from Patent EP1288294.  
ACCESSION AX709919  
VERSION AX709919.1 GI:29786301  
SOURCE Candida rugosa  
ORGANISM Candida rugosa  
Bukariyota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE  
1  
AUTHORS Tang, S.J., Lee, G.C. and Shaw, J.F.  
TITLE Recombinant candida rugosa lipases  
JOURNAL Patent: EP 1288294-A 1 05-MAR-2003;  
Academia Sinica (TW)  
FEATURES  
Location/Qualifiers  
1..1469  
/organism="Candida rugosa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:5481"

ORIGIN  
Query Match 54.2%; Score 830.4; DB 6; Length 1469;  
Best Local Similarity 78.1%; Pred. No. 3e-127;  
Matches 1202; Conservative 0; Mismatches 241; Indels 97; Gaps 13;  
QY 10 CAACTCGGCAAGCGGACACCATCACCGTCTCAACGCGATCATCAACGAGCGTTCCT 69  
Db 8 CCACCTCGCAACGCGGACACCATCACCGTCTCAACGCGATTTGTCAACGAAAGTTCT 67  
QY 70 CGGCATTCCTTTTGGCGAGCGCGGTGGGCAACCTCCGCTTCAAGGACCTGTGCGGTA 129  
Db 68 CGGCATACGTTTGGCGAGCGCGCGGTGGCA--CTCCGCTTCAA-----110  
QY 130 CTCTGCTCGCTCAACGCGCAGAAG-TTACTTACGCGCGGTGCATGCGAGAGAACCCGA 188  
Db 111 -----GCTCAACGCGCAGAGTTTACTACGCGCGGTGCATGAGATGAACCTAT 161  
QY 189 GGGCAGTTTGAAGAGAACCTTGGCAAGCGCACTCGACTTGGTGTGATGCTCAGTCCAGT 248  
Db 162 GGGCTCGTTTCA-----TTGGTGTCTCAGTCCAAGAT 193  
QY 249 GTTCCAGGCGGTGCTTCCCGAGTGAGGACTGCTCACCATCAAGTGGTGGCGCGCC 308  
Db 194 CTTCAAGTGTGTTTCCCAACGAGGACTGTCTCAC-----232  
QY 309 GGGCACCAAGGCGGCGGCAACCTCCCGGTCTGCTTGGATCTTTGGCGGTGGTTGA 368  
Db 233 ----CACCGGCGAGTGTGCTCCCGGTGATGCTCTGATCTTTGGCGGTGGTTGA 289  
QY 369 GATCGGAGCGCCCAACATCTTCCCTCCCGCCAGATGTGTCAACAAGAGTGTCTATGG 428  
Db 290 GCTTGGCGCTCCAGCTCTTTCCAGGAGACAGATGGTGGCGCAAGAGCGTGTCTATGG 349

Qy	1493	CGTGGCTACGACGCGTTGATGACCAACCCGCTTCTTTGTG	1532
Db	1428	GGATGCGTACAGCGCCCTCTTTTCCAAACCGCCTTCTTTG	1467
RESULT 15			
AF044078		1688 bp DNA linear	SYN 07-FEB-1998
LOCUS			
DEFINITION		Synthetic construct triacylglycerol hydrolase (lip1) gene, complete	
ACCESSION	AF044078		
VERSION	AF044078.1	GI:2852389	
KEYWORDS			
SOURCE		synthetic construct	
ORGANISM		artificial sequences.	
REFERENCE		1 (bases 1 to 1688)	
AUTHORS		Brocca, S., Schmidt-Dannert, C., Lotti, M., Alberghina, L. and Schmid, R.D.	
TITLE		Design, total synthesis and functional overexpression of the Candida rugosa lip1 gene coding for a major industrial lipase	
JOURNAL		Protein Sci. (1998) In press	
REFERENCE		2 (bases 1 to 1688)	
AUTHORS		Brocca, S., Schmidt-Dannert, C., Lotti, M., Alberghina, L. and Schmid, R.D.	
TITLE		Direct Submission	
JOURNAL		Submitted (22-JAN-1998) Institute of Technical Biochem., University of Stuttgart, Allmandring 31, Stuttgart 70569, Germany	
FEATURES		Location/Qualifiers	
source		1..1688	
		/organism="synthetic construct"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:32630"	
gene		/note="designed gene encoding lip1 of Candida rugosa"	
		19..1688	
		/gene="lip1"	
CDS		19..1688	
		/EC_number="3.1.1.3"	
		/function="lipase"	
		/note="Lip1; codon optimized for expression in Saccharomyces cerevisiae"	
		/codon_start=1	
		/product="triacylglycerol hydrolase"	
		/protein_id="AAC02093.1"	
		/db_xref="GI:2852390"	
		/translation="MELALSLIASVAAPTATLANGDTITGLNAILNEAPLGIIPFA EPPGNLRFDPVPYSGLDQKFTSYGSPCMQONPBGTYEENLPKALDLVMSKVF EAVSPSEDCLTINNVPPPGTKAGANLPVLMWIFGGFEVGTSTFPFPAQMITSIAM KPTIHVSVYRVSSWGLAGDEIKAGSANAGLKQRLGMQWADVDAVGIYGNETFDL TIPESAGSMVMCHILWNDGNTYKGLPLFRAGIMOSGAMVPSDVPDGVNIITDDM LASNAGCGSADKLACLRGVSSDLEADNTNTPGLAYSLRLSYLPRPDGVNITDDM YALVREGKYANIPIVIGDONDEGTFEGTSSILNVTDAQAREVKOSFVHASDABIDTL MTATPGDITQSGPFDTGILNALTQFRIKSAVLGDLGFTLARRVFLNHYTGGTKYSEL SKQUSGLPVGTFHSDIVFDYLLSGSLIYNNAFAFADLDPDTPAGLLVKKPEYIT SSSQSGNNLMINALGLYTGKDNFRTAGYDALFSPNPPFFV"	
ORIGIN			
Query Match		42.5%; Score 651.8; DB 12; Length 1688;	
Best Local Similarity		69.7%; Pred. No. 8.2e-98;	
Matches 1104; Conservative		0; Mismatches 417; Indels 62; Gaps 14;	
Qy	2	CCCACCGCAAGCTCGCCACGCGGACACCATCACCGGTCTCAACGCCATCATCAACGAG	61
Db	67	CCACCGCCACCTTTGGCTAACGGTGACACCATCACCGGTTTGAACGCCATCATCAACGA	126
Qy	62	CGCTTCTCGGCATTCCTTTTCCGAGCGCGCGGTGGCAACCTCCGCTTCAAGACCT	121
Db	127	GCCTTCTGGGTATTCCATTTCGCCGACACCACTGGTGTGTAATTCAGTTCAGGACCA	186
Qy	122	GTGCGGTACTCTGCTCGCTCAACGGCCAGAGTTACTTACGGCC-----CGTGCATG	174
Db	187	GTTCCTACTCCGGTTCCTTGGATGGTCAAAAGTTCACTTCTTAGGGTCCATCTTGTATG	246

Qy	175	CAGCAGAACCCGAGGSCAGTTTGAAGAGAACCTTGGCAAGACGGCACTCGACTTGGTG	234
Db	247	CAACAAAACCCAGAGGTACTACGAAGAAATCTGCCAAGGAGCTTTAGATCTGGTT	306
Qy	235	ATGCAGTCCAGAGTGTTCAGGCGGTCTTCCCGAGAGTGAAGACTGCTCACCATCAAC	294
Db	307	ATGCAATCCAAAGTTTTCGAAGCTGTCTCCATCTCTTGAAGACTGTTTGACCATTAAT	366
Qy	295	GTGTTGGCGCGCGGACCAAGGGCGGCGCAACCTCCGGTCAATGCTCTGATCTTT	354
Db	367	GTGTTAGACACCCGGGCAAGGCTGTGGCCAACTTGCAGTATGTGTGATCTTT	426
Qy	355	GGCGGTGGGTTTGAGATCGGAGCCCAACCATCTTCCCTCCCGCCAGATGGTCACCAAG	414
Db	427	GGTGGTGGTTTGAAGTGGTGGTACTAGTACCTTCCCTCCAGCCCAATGATACCAAG	486
Qy	415	AGTGTGCTCATGGGCAGCC-ATCATCCAGTGGCCGTCAACTACCGTGTGCTCGTGG	473
Db	487	TCTATTGCTATGGGTAAGCCCAATCATCCAGCTTTCTGTCAACTACAGAGTCTCGAGCTGG	546
Qy	474	GGGTTCTTGGCTGTGATGACATCAAGGCCGAGGCGAGCGGAAACCGCGCTTGAAGGAC	533
Db	547	GGTTCTTGGCTGGTGAAGAAATCAAGCCGAGAGTTCTGCCAAGCCGGTTTGAAGGAC	606
Qy	534	CAGCGTTTGGGCATGCAATCGGTTGGCAGACAAATTCGCCGGTTTGGCGGCGACCCGAGC	593
Db	607	CAAGATTTGGGTATGCAATGGTGGTGGTGACAAACATTCGCTTTTGGTGGTATCCAACT	666
Qy	594	AAGGTGAC-ATCTTTGGCGA---GGCGGGCAGCATGTCGCTGTGTGGCCACCTCATCTGG	649
Db	667	AAGTTACTATCTTTGGTGAATCTGCTGGTTCTATGTCCGTCATGTGTGCACATTTTGTGG	726
Qy	650	AACGAGGCGGACAAACAGTACAAGGCAAGCCGTTGTTCCGCGCGGCATCATGCA---G	706
Db	727	AACGAGGTGACAACTTACAGGGTGAAGCAATGTTTACAGCTGGTATCATGCAATCT	786
Qy	707	GGAGCCATGTGC---CGGACCCGGTGGAGCGGACGCTACGGCAACGAGATCTACGACCTC	763
Db	787	GGTGTATGGTTCCATCTGACGCGCTGACGAGTATCTACGGTAACGAAATTTTGTACTTG	846
Qy	764	TTTGTCTCGAGTGTGCTGTGCGAGGCGGAGGCAAGCTCGGTGCTTGGCGAGTG--	821
Db	847	TTGGCTTCCAAACGCTGGTGTGGTTCTGCTCTGACAAAGTTGGCTTGTGGAGAGTGT	906
Qy	822	-CGAGCGACACCTTGCTCGATGCCAACAAACACTCTCGGTTCTTGGGTACTCTCG	880
Db	907	TCTTCTGACACTTTGGAAGAGCGCAACCAACACACCCCTGTTCTTGGCTTACTCTCTC	966
Qy	881	TTGCGGTTGTAAT-----CCGCGCGCGGAGGCAAGCAATCACCGATGACATGTACAG	934
Db	967	TTAAGATTGTCTTACTTGGCAAGACGAGCGGTGTTAAACATCACCGACGACATGTACGCT	1026
Qy	935	TTGGTGGCGAGCGCAAGTATGAAGGTTCCGCTGATCATTTGGCGACGAGACGAGAG	994
Db	1027	TTGGTTAGAGAGGTAAATGATGCAACATCCCTGTTATCATCGGTGACCAAAACGACGAA	1086
Qy	995	GGCACCATCTTTGGCTC-----TTGAAGTGCACCAAGTGTCTCAGGCCGCTGTAC	1047
Db	1087	GGTACCTTCTTGGTACTTCTTTTGAAGCTTACCACTGATGCCAAGCCAGAGAAAT	1146
Qy	1048	TTCAAGCA---GTTTATCCAGCGAGCGAGATCGACACCTTTGATGGCGGGTAC	1104
Db	1147	TTCAAGCAATCTTTTGTCCAGCTAGCAGCTGAAATCGACACTTTGATGATGCTGTAC	1206
Qy	1105	CCCCAGACATCAACCCAGGT-----CGTTTCGACACCGGTTCAAGGTCACCCCG	1154
Db	1207	CCAGGTGACATCACTCAAGGTTCTCCATTTGACACTGGAATCTAAACGCTTGAACCCCA	1266
Qy	1155	CAGTTCAAGAGATC---GGGTGCTCGGAGCTTGCATTCATCCACGCCCGCGCTAC	1211
Db	1267	CAATTCAAGAGAACTCTGCTGTTTGGGTGACTTGGGTTTACTTTGGCTCGATATAC	1326

Search completed: August 5, 2004, 17:10:02  
Job time : 4058 secs

Qy	1212	TTCTCAACCACTTCCAGGGCGGCACCAAGTACTCGTTCTCAAGCAGCTC-----GGG	1265
Db	1327	TTCTTGAACCACTACACCGGTGGTACCAAGTACTCTTTCTTGTCTAAGCAATTTCTTGGT	1386
Qy	1266	TTGCAATCATGGGCACTTTCCATGCAACGACATTTGTGTGGCAGGACTACTTTGTGGGA	1325
Db	1387	TTGCCAGTTTGGGTACTTTCCACTCCACGATATCGTCTTCCAAAGACTACTTTGTGGGT	1446
Qy	1326	AGCGGAGCGTCTACTACAAACGCGTTTATCGGTTCCGCCACCGACTTGGACCCCAAC	1385
Db	1447	TCTGGTTCTTGTATCTACAAACAGCTTTTCAATGCTTTTGGCACTTGGACCCCAAC	1506
Qy	1386	ACGCGGGGTTGTTGGTGAATCTGGCCCAAGTACACC-----AGCAGCCAGGGCAACAAC	1439
Db	1507	ACGCGCGTTTGTGTTAAGTGGCCAGAAATACACCTCTTCTCTCAATCTGGTAAACAAC	1566
Qy	1440	TTGATGATGATCAACCGCTTGGGTTGTACACCGGCAAGGACAACTTCCGACCGCTGGC	1499
Db	1567	TTGATGATGATCAACCGCTTGGGTTGTACACCGGTAAGGACAACTTTCAGAACCGCGGT	1626
Qy	1500	TACGACGCGTTGATGACCAACCC	1522
Db	1627	TACGACGCTTGTCTTCCAAACCC	1649

B/gnK



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2004, 14:49:49 ; Search time 466 Seconds  
(without alignments)  
13966.179 Million cell updates/sec

Title: US-09-943-857-3

Perfect score: 1532

Sequence: 1 gccaccgcgaagctcgcca.....tgaccacccgtttctgtg 1532

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1268	82.8	1641	7 ABX95906	Abx95906 Candida r
2	1020.4	66.6	1650	2 AAT10422	Aat10422 Candida c
3	1018.4	66.5	1641	7 ABX95909	Abx95909 Candida r
4	991.6	64.7	1649	2 AAX33111	Aax33111 Candida r
5	968.6	63.2	1641	7 ABX95908	Abx95908 Candida r
6	932	60.8	1641	7 ABX95907	Abx95907 Candida r
7	873.2	57.0	1641	7 ABX95905	Abx95905 Candida r
8	651.8	42.5	1647	2 AAX33112	Aax33112 Synthetic
9	651.8	42.5	1950	4 AAH43625	Aah43625 Synthetic
10	173	11.3	2045	2 AAQ98578	Aaq98578 Aspergill
11	173	11.3	2045	2 AAT11299	Aat11299 Aspergill
12	171.2	11.2	1674	2 AAQ05605	Aaq05605 Gene enco
13	162	10.6	1828	2 AAQ54020	Aaq54020 Lipase co
14	160.2	10.5	1635	2 AAT10421	Aat10421 Geotrichu
15	157.8	10.3	1692	2 AAQ10313	Aaq10313 Sequence
16	88.2	5.8	1716	4 AAH26618	Aah26618 Rhodospo
17	88.2	5.8	1738	2 AAQ21460	Aaq21460 Rhodospo
18	87	5.7	8528	2 AAQ46249	Aaq46249 Construct
19	62.8	4.1	2428	2 AAQ83221	Aaq83221 Bile salt
20	61.2	4.0	2184	2 AAO71874	Aao71874 Sequence
21	61.2	4.0	2312	6 ABQ54978	Abq54978 Human ova
22	61.2	4.0	2428	2 AAQ20452	Aaq20452 Bile-salt
23	61.2	4.0	2428	2 AAT35250	Aat35250 Human bil

24	61.2	4.0	2428	2	AAZ20300	Aaz20300 Human bil
25	61.2	4.0	2428	6	ABL70012	Abi70012 Pancreas
26	61.2	4.0	2428	6	ABA04403	Abi04403 Human mil
27	61.2	4.0	2428	6	AAS18538	Aas18538 cDNA enco
28	61.2	4.0	2487	3	AAA63883	Aaa63883 cDNA enco
29	60.6	4.0	2000	7	ADA71938	Ada71938 Rice gene
30	60	3.9	1725	3	ABA97180	Aba97180 Synthetic
31	60	3.9	1752	6	AAS17549	Aas17549 Ache codi
32	60	3.9	1845	6	AAS17493	Aas17493 Human cdn
33	60	3.9	1936	7	ACC47509	Acc47509 Human ace
34	60	3.9	2156	7	ABZ23128	Abz23128 Nucleotid
35	60	3.9	2256	2	AAQ99002	Aaq99002 Human ace
36	60	3.9	2567	6	AAS17547	Aas17547 Plasmid p
37	60	3.9	12113	6	AAS17492	Aas17492 Human ace
38	60	3.9	14446	6	AAS17548	Aas17548 Plasmid p
39	59.6	3.9	1614	2	AAQ90569	Aaq90569 Recombina
40	59.6	3.9	2166	2	AAT47862	Aat47862 Human wil
41	59.6	3.9	2166	2	AAQ90579	Aaq90579 Encodes b
42	59.6	3.9	2166	2	AAT47861	Aat47861 cDNA of h
43	59.6	3.9	2344	4	AAF31110	Aaf31110 Cholesterol
44	59.6	3.9	3004	6	ABL62109	Abi62109 Colon ade
45	59.6	3.9	3018	2	AAQ14186	Aaq14186 Human mil

## ALIGNMENTS

RESULT 1  
ABX95906  
ID ABX95906 standard; DNA; 1641 BP.  
XX  
AC ABX95906;  
XX  
DT 15-JUL-2003 (first entry)  
XX  
DE Candida rugosa lipase 3 DNA.  
XX  
KW Lipase 3; gene; ds; mutant.  
XX  
OS Candida rugosa.  
XX  
FH Key  
CDS 1. .1641  
FT /\*tag= a  
FT /product= "Lipase 3"  
FT /partial  
FT /note= "No start or stop codon shown"  
XX  
PN EP1288294-A2.  
XX  
PD 05-MAR-2003.  
XX  
PF 26-APR-2002; 2002EP-00009616.  
XX  
PR 31-AUG-2001; 2001US-00943857.  
XX  
PA (SINI-) ACAD SINICA.  
XX  
PI Tang S, Lee G, Shaw J;  
XX  
DR WPI; 2003-395476/38.  
XX  
PS P-PSDB; ABU09071.  
XX  
PT Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for  
PT the preparation of Candida rugosa lipase for biocatalytic applications.  
XX  
PS Claim 25; Page 5-7; 33pp; English.  
XX  
CC The invention relates to an isolated mutant nucleic acid encoding a  
CC Candida rugosa lipase polypeptide. The DNA has a sequence having at least  
CC 80% identity to a wild-type DNA encoding Candida rugosa lipase. The  
CC invention also relates to a microorganism comprising the DNA, where the  
CC microorganism is a bacterium or yeast, preparing a mutant DNA encoding a

CC C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate  
interacting domain of a first C. rugosa lipase and a non-substrate  
interacting domain of a second C. rugosa lipase. The method is useful for  
preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is  
useful in the large scale manufacture of Candida rugosa lipase which is  
useful for biocatalytic applications. This sequence represents DNA  
encoding Candida rugosa lipase 3

XX  
SQ Sequence 1641 BP; 315 A; 525 C; 459 G; 342 T; 0 U; 0 Other;

Query Match 82.8%; Score 1268; DB 7; Length 1641;  
Best Local Similarity 94.5%; Pred. No. 1.3e-266;  
Matches 1505; Conservative 0; Mismatches 25; Indels 62; Gaps 16;

QY 2 CCCACCGCAGCTCGCCCAACGGGACACCATCAACCGGTCTCAACGGCCATCATCAACGAG 61  
DB 43 CCCACCGCAGCTCGCCCAACGGGACACCATCAACCGGTCTCAACGGCCATCATCAACGAG 102  
QY 62 GCGTTCCTCGGCATTTCCCTTTGCGAGCGCGCGGTGGGCAACCTCCGCTTCAAGGACCT 121  
DB 103 GCGTTCCTCGGCATTTCCCTTTGCGAGCGCGGTGGGCAACCTCCGCTTCAAGGACCT 162  
QY 122 GTGCGGTACTCTGGCTCGCTCAACGGCCAGAGT-----TACTTACGGCCCG---TGCA 174  
DB 163 GTGCGGTACTCTGGCTCGCTCAACGGCCAGAGTTCACCTTCTTACGGCCCGCTTTCGATG 222  
QY 175 CAGCAGAACCCCGAGGGGACGTTTGAAGAGAACCTTGGCAAGCGGCATCGACTTGGTG 234  
DB 223 CAGCAGAACCCCGAGGGGACGTTTGAAGAGAACCTTGGCAAGCGGCATCGACTTGGTG 282  
QY 235 ATGAGTCCAAAGTGTCTCAGCGGTGCTTCCCAAGAGTGGAGTGCCTCACCATCAAC 294  
DB 283 ATGAGTCCAAAGTGTCTCAGCGGTGCTTCCCAAGAGTGGAGTGCCTCACCATCAAC 342  
QY 295 GTGCTGCGCGCGCGCGGACCAAGCGCGCGCGCAACCTCCCGGTCACTGCTCGATCTTT 354  
DB 343 GTGCTGCGCGCGCGCGGACCAAGCGCGCGCGCAACCTCCCGGTCACTGCTCGATCTTT 402  
QY 355 GCGGTGCGGTGAGATGGGAGCGCGCACCATCTTCCCTCCCGCGCGCATGTGTCACCAAG 414  
DB 403 GCGGTGCGGTGAGATGGGAGCGCGCACCATCTTCCCTCCCGCGCGCATGTGTCACCAAG 462  
QY 415 AGTGTGCTCATGGGCAAGC-CATCATCCAGTGGCGGTCACTTACCGTGTTCCTTCGTGG 473  
DB 463 AGTGTGCTCATGGGCAAGC-CATCATCCAGTGGCGGTCACTTACCGTGTTCCTTCGTGG 522  
QY 474 GGGTCTTGGCTGCTGATGACATCAAGCGCGGAGCGGAAACCGCGGTGTAAGGAC 533  
DB 523 GGGTCTTGGCTGCTGATGACATCAAGCGCGGAGCGGAAACCGCGGTGTAAGGAC 582  
QY 534 CAGCGTTTGGGATGCGAGTGGGTGGCAGACACATTCGCGGTTCGCGCGGACCGGAGC 593  
DB 583 CAGCGTTTGGGATGCGAGTGGGTGGCAGACACATTCGCGGTTCGCGCGGACCGGAGC 642  
QY 594 AAGGTGAC-ATCTTTGGCGAG--GCGGACAGTGTCCGTGTGTGGCCACCTCATCTGG 649  
DB 643 AAGGTGACATCTTTGGCGAGTGTGGGCGAGCATGTCCGTGTGTGGCCACCTCATCTGG 702  
QY 650 AACACGGCGGACAAACGTAAGGCGGAGCGCGGTGTTCCGCGCGGCGATCATGCG--- 706  
DB 703 AACACGGCGGACAAACGTAAGGCGGAGCGCGGTGTTCCGCGCGGCGATCATGCGAGTCT 762  
QY 707 GGAGCCATGTTGCGG---GACCGGTGGAGCGGACGTAGGCAACGAGATCTACGACCTC 763  
DB 763 GGAGCCATGTTGCGGCTGACCGCGGTGGAGCGGACGTAGGCAACGAGATCTACGACCTC 822  
QY 764 TTTGTCTCCAGTGTCTGCTGTGGCAGCGCCAGCAGCAAGTCCGCTGTGGCAGTGGG 823  
DB 823 TTTGTCTCCAGTGTCTGCTGTGGCAGCGCCAGCAGCAAGTCCGCTGTGGCAGTGGG 882  
QY 824 ---AGCGACACCTTGTCTGATGCGCAACAAACACTCTCTGGGTCTTGGGTACTCTCG 880  
DB 883 TCTAGCGACACCTTGTCTGATGCGCAACAAACACTCTCTGGGTCTTGGGTACTCTCTCG 942

QY 881 TTGCGGTGTG-----ACTCCCGCGCGCGGCAAGCAATCAACGATGATGATCAAG 934  
DB 943 TTGCGGTGTGTTATCTCTCCCGCGCGCGGCAAGCAATCAACGATGATGATCAAG 1002  
QY 935 TTGCGTGGCGACCGGAGTATGCAAGGTTCCCGTGTATCATTTGGCGACCAAGACGAG 994  
DB 1003 TTGCGTGGCGACCGGAGTATGCAAGGTTTCCCGTGTATCATTTGGCGACCAAGACGAG 1062  
QY 995 GGACCATCTTTGG-----CTCTTTGAACGTGACCAAGTGTCTCAGGCGCGGTCTTAC 1047  
DB 1063 GGACCATCTTTGGGCTCTCTTTTGAACGTGACCAAGTGTCTCAGGCGCGGTCTTAC 1122  
QY 1048 TTCAAGCAG---TTTATTCACCGCAGCGGAGATCGACACTTGTATGGCGCGGTAC 1104  
DB 1123 TTCAAGCAGTCTTTTATTCACCGCAGCGGAGATCGACACTTGTATGGCGCGGTAC 1182  
QY 1105 CCCGAGACATCAACCGAGG---TCCGTTGACACACGTT-----CAAAGCTCACCCCG 1154  
DB 1183 CCCGAGACATCAACCGAGGTTTCTCGTTTGAACACGAGTGTCTCAACGCGCTCACCCCG 1242  
QY 1155 CAGTTCAAGAGATC---GCGGTGCTCGGCACTTGCATTCATCCAGCGCGCGGTAC 1211  
DB 1243 CAGTTCAAGAGATCTCTCGGTGCTCGGCACTTGCATTCATCCAGCGCGCGGTAC 1302  
QY 1212 TTCCTCAACCACTTCCAGGGCGGCAACAGTACTCGTTCCTC-----AAGCAGTCTGG 1265  
DB 1303 TTCCTCAACCACTTCCAGGGCGGCAACAGTACTCGTTCCTCTTAAGCAGTCTCTGG 1362  
QY 1266 TTGCAATCATGGGACCTTCCATGCAACGACATTTGTGGCAGGACTACTTGTGGGA 1325  
DB 1363 TTGCAATCATGGGACCTTCCATGCAACGACATTTGTGGCAGGACTACTTGTGGGA 1422  
QY 1326 AGCGGACGCTCATCTCAACACGCGTTTATCGGTTTCGCCACGACTTGGACCCCAAC 1385  
DB 1423 AGCGGACGCTCATCTCAACACGCGTTTATCGGTTTCGCCACGACTTGGACCCCAAC 1482  
QY 1386 ACCCGGGGTTTGGTGAACTGGCCCAAGTATACACGACG-----CAGGGCAACAC 1439  
DB 1483 ACCCGGGGTTTGGTGAACTGGCCCAAGTATACACGACG-----CAGGGCAACAC 1542  
QY 1440 TTGATGATCATACGCGCTTGGGCTTGTACACCGCAAGGACAACTTCCGACCGCTGGC 1499  
DB 1543 TTGATGATCATACGCGCTTGGGCTTGTACACCGCAAGGACAACTTCCGACCGCTGGC 1602  
QY 1500 TACGACGCTGATGACCAACCGCTTCTTTGT 1531  
DB 1603 TACGACGCTGATGACCAACCGCTTCTTT 1634

RESULT 2

AAAT10422  
ID AAAT10422 standard; DNA; 1650 BP.  
XX  
AC AAAT10422;  
XX  
XX 20-SEP-1996 (first entry)  
XX  
XX Candida cylindracea lipase gene.  
DE  
XX  
KW Fatty acid; transgenic plant; exogenous; lipase; Rhizopus niveus;  
KW Pseudomonas aeruginosa; Pseudomonas fluorescens; Geotrichum candidum;  
KW Candida cylindracea; constitutive; tissue specific; promoter; lipid;  
KW milling; biofuel; lubricant; detergent; ss.  
OS  
XX Candida cylindracea.  
XX  
XX Key Location/Qualifiers  
FH CDS 46..1647  
FT /\*tag= a  
FT /product= "lipase enzyme"  
XX  
XX  
PN FR272798-A1.

XX	26-JAN-1996.	Db	589	CAGCGCTTGGCATGTCAGTGGTGGCGGACAAATTCGGCGGCTTTGGCGGACCCCGACC	648
XX	25-JUL-1994; 94FR-00009272.	Qy	594	AAGTGA-CATCTTTGGCGA---GGCGGAGAGATGTCGGTGTGGTGGCCACTCATCTGG	649
XX	25-JUL-1994; 94FR-00009272.	Db	649	AAGTGAACCATCTTTGGCGAGCTGGCGGAGAGATGTCGGTGTGGTGGCCACTCATCTGG	708
XX	(NAPO-) INST NAT POLYTECHNIQUE TOULOUSE.	Qy	650	AAGCGCGGACAAACGTAACAGGGAACCGTGTTCGGCGGCGGCATCATGCA---G	706
XX	Alibert G, Mouloungui Z, Boudet A;	Db	709	AACGACGGCGACAAACGTAACAGGGAACCGTGTTCGGCGGCGGCATCATGCACTG	768
XX	WPI; 1996-107680/12.	Qy	707	GGAGCCATGCTGCC---GGAACCGCGTGGACGGCAGCTACGGCAACGAGATCTACGACCTC	763
XX	Prodn. of fatty acids or derivs. from transgenic oilseed plants -	Db	769	GGGCGCATGCTGCGCTGGACGCGGTGACGATCTACGGCAACGAGATCTTTGACCTC	828
XX	engineered to express a lipase that contacts lipid(s) only when seeds are	Qy	764	TTTGTCTCAGTGTGCTGGTGTGGCAGGCGCAGGCAAGCTCGCTGCTTGGCAGTGC-	822
XX	milled.	Db	829	TTGGCGTGAACCGCGGCTGCGGACGGCGGCAAGCTTGCCTGCTTGGCGGCTGTG	888
XX	Claim 9; Page 26-27; 32pp; French.	Qy	823	--GAGCGACACCTTGTCTGATGCCAACCAACACTCTCTGGTCTTTGGCGTACTCCTCG	880
XX	A novel method of producing fatty acids or their derivs. in a plant	Db	889	CTGAGCGACACGTTGGAGGAGCGCAACAAACACCCCTGGGTCTTTGGGTACTCTCCTCG	948
XX	comprises generating a transgenic plant contg. an exogenous lipase gene,	Qy	881	TTGCGGTG---TACTTCCGCGCGGACGCAAGAACATCACCGATGACATGTACAAG	934
XX	esp. selected from the <i>Rhizopus niveus</i> , <i>Pseudomonas aeruginosa</i> ,	Db	949	TTGCGGTGCTGTGTACCTCCCGCGCGGCGGTGAACATCACCGACGATGTAGGCC	1008
XX	<i>P. fluorescens</i> , <i>Pseudomonas</i> sp., <i>Geotrichum candidum</i> or <i>Candida</i>	Qy	935	TTGGTGGCGGACGGCAAGTATGCAAGGTTCCCGTGTATCTTGGCGACGAGACGAG	994
XX	<i>cylindracea</i> lipase genes (AAT10417-22 resp.). The lipase genes can be	Db	1009	TTGGTGGCGGAGGGCAAGTATGCCAATCCCTGTGTATCATCGCGACCAAGACGAG	1068
XX	placed under control of a constitutive or a tissue specific promoter. The	Qy	995	GGCACCATCTTTG---GCTCTTGAACGTCGACCAACAACTCTCAGGCGCGTCTTAC	1047
XX	production of the fatty acids only occurs when the lipase and lipids	Db	1069	GGCACCTCTTTGGGACCCCTGCTGTGTAAGTGAACGATGCCAGGCGCCGAGTAC	1128
XX	contact each other after milling of the plants. The fatty acids generated	Qy	1048	TTCAAGCA---GTTTATCAACGCGACGCGAGATCGACACTTGTATGGCGGGTAC	1104
XX	can be used to prod. e.g. biofuels, lubricants, detergents, etc	Db	1129	TTCAAGCAGCTGTTTGTCCAGCGCAGCGCGGAGATCGACAGTTGATGACGCGTAC	1188
XX	Sequence 1650 BP; 299 A; 540 C; 491 G; 320 T; 0 U; 0 Other;	Qy	1105	CCCCAGGACATCACCCAGGG---TCCGTTGACACGCGTT-----CAAGCTCACCCCG	1154
XX	Query Match 66.8%; Score 1020.4; DB 2; Length 1650;	Db	1189	CCCGCGCATGATGACCCAGGCGCTGCGGTTTCGACACGGGTATTCTCAACGCGCTCACCCCG	1248
XX	Best Local Similarity 84.7%; Pred. No. 1.2e-212;	Qy	1155	CAGTTCAAGAGAAATC---GCGGTGCTCGCGACCTTGTGATTCATCCACGCGCGCTAC	1211
XX	Matches 1341; Conservative 0; Mismatches 181; Indels 62; Gaps 15;	Db	1249	CAGTTCAAGAGAAATCCTCGCGGTGCTCGCGGACCTTGGCTTTACGCTTGTCTCGCTAC	1308
Qy	2 CCACCGCAAGCTCCGACGCGGACCAACATCACCGTCTCAACGCCATCATCAACGAG	Qy	1212	TTCTCAACCACTTCAGGGCGGACCAAGTACTCGTTCTC-----AAGCAGCTCGGG	1265
Db	49 CCACCGCACGCTCCGACCGGACCAACATCACCGTCTCAACGCCATCATCAACGAG	Db	1309	TTCTCAACCACTACACCGCGGACCAAGTACTCATTCCTCTCTGAAGCAGCTCCTGGGC	1368
Qy	62 GGTTCCTCGGATTCCTTTGCGGAGCGCGGTGGCAACCTCGCTTCAGGACCTT	Qy	1266	TTGCCAATCATGGGACCTTCCATGCCAACGACATGTGTGGCAGGACTACTTTTGGGA	1325
Db	109 GGTTCCTCGGATTCCTTTGCGGAGCGCGGTGGCAACCTCGCTTCAGGACCTT	Db	1369	TTGCCGCTGCTCGGAACGTTCCACTCCAAACGACATTTGCTTCAGGACTACTTTGGGC	1428
Qy	122 GTGCGGATCTGCGCTCGCTCAACGCCGAGAGTT-----ACTTACGCGCGCTGCATG	Qy	1326	ACGCGCAGCGCTCATCTACAAACAAACCGTTTATCGGTTTCGCCACGACTTGGACCCCAAC	1385
Db	169 GTGCGGATCTGCGCTCGCTCAACGCCGAGAGTTCAACGCCGCTGCATG	Db	1429	ACGCGCTCGCTCATCTACAAACAAACCGTTTATCGGTTTCGCCACGACTTGGACCCCAAC	1488
Qy	175 CAGCAGAACCCGAGGCGACGTTTGAAGAGACCTTGGCAAGCGACTCGACTTGGTG	Qy	1386	ACCGCGGGTGTGTGGTGAACCTGGCCCAAGTACACAGCAGC-----CAGGGCAACAC	1439
Db	229 CAGCAGAACCCGAGGCGACCTTGAAGAGACCTTGGCAAGCGACTCGACTTGGTG	Db	1489	ACCGCGGGTGTGTGGTGAACCTGGCCCAAGTACACAGCAGC-----CAGGGCAACAC	1548
Qy	235 ATGCAATCCAAAGTGTTCAGGCGGTGCTTCCAGAGTGAGGACTGCTCACCATCAAC	Qy	1440	TTGATGATGATCAACGCGCTTGGGCTTGTACCCGCGAGGACAACTTCGCAACCGCTGGC	1499
Db	289 ATGCAATCCAAAGTGTTCAGGCGGTGCTTCCAGAGTGAGGACTGCTCACCATCAAC	Db	1549	TTGATGATGATCAACGCGCTTGGGCTTGTACCCGCGAGGACAACTTCGCAACCGCTGGC	1608
Qy	295 GTGGTGGCGCGCGGACCAAGCGGCGGCAACCTCCCGGTGATGCTCTGGATCTTT	Qy	1500	TACGACGCGTTGATGACCAACCCG	1523
Db	349 GTGGTGGCGCGCGGACCAAGCGGCGGCAACCTCCCGGTGATGCTCTGGATCTTT	Db	1609	TACGACGCGTTGTTCTCCACCCG	1632
Qy	355 GCGGTTGGTGTGAGATCGGAGGCCCAACATCTTCCCTCCGCGCAGATGTACCAACAG	Qy			
Db	409 GCGGCGGTTGTGAGTGGTGGCGACCAACGACTTCCCTCCGCGCAGATGATCAACAG	Db			
Qy	415 AGTGTGCTCATGGGCAAG-CCATCATCCACGTGGCGTCAACTACCGTGTTCGCTCGTG	Qy			
Db	469 ASCATGTCATGGCAAGCCCATCATCCAGTGAAGTCACTACCGGTGCTGCTGGG	Db			
Qy	474 GGGTCTTGGCTGTGATGATCAAGCGCGGAGGAGGAAACCGCGCTTGAAGGAC	Qy			
Db	529 GGGTCTTGGCTGTGATGATCAAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	Db			
Qy	534 CAGCGTTTGGGATGAGTGGTGGCAGAACATTCGCGGTTTCGCGGCGACCCGAGC	Qy			

ID ABX95909 standard; DNA; 1641 BP.  
 XX AC ABX95909;  
 XX DT 15-JUL-2003 (first entry)  
 XX DE Candida rugosa lipase 8 DNA.  
 XX KW Lipase 8; gene; ds; mutant.  
 XX OS Candida rugosa.  
 XX PH Key Location/Qualifiers  
 FT CDS 1. 1641  
 FT /\*tag= a  
 FT /product= "Lipase 8"  
 FT /partial  
 FT /note= "No start or stop codon shown"  
 XX  
 PN EPI288294-A2.  
 XX  
 XX 05-MAR-2003.  
 XX PD 26-APR-2002; 2002EP-00009616.  
 XX PF 31-AUG-2001; 2001US-00943857.  
 XX PR (SINI-) ACAD SINICA.  
 XX PA Tang S, Lee G, Shaw J;  
 XX PI WPI; 2003-395476/38.  
 XX DR P-PSDB; ABU09074.  
 XX  
 PT Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for  
 PT the preparation of Candida rugosa lipase for biocatalytic applications.  
 XX  
 PS Claim 25; Page 10-11; 33pp; English.  
 XX  
 CC The invention relates to an isolated mutant nucleic acid encoding a  
 CC Candida rugosa lipase polypeptide. The DNA has a sequence having at least  
 CC 80% identity to a wild-type DNA encoding Candida rugosa lipase. The  
 CC invention also relates to a microorganism comprising the DNA, where the  
 CC microorganism is a bacterium or yeast, preparing a mutant DNA encoding a  
 CC C.rugosa lipase and a chimeric C.rugosa lipase comprising a substrate  
 CC interacting domain of a first C.rugosa lipase and a non-substrate  
 CC interacting domain of a second C.rugosa lipase. The method is useful for  
 CC preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is  
 CC useful in the large scale manufacture of Candida rugosa lipase which is  
 CC useful for biocatalytic applications. This sequence represents DNA  
 CC encoding Candida rugosa lipase 8  
 XX  
 SQ Sequence 1641 BP; 298 A; 537 C; 472 G; 334 T; 0 U; 0 Other;  
 Query Match 66.5%; Score 1018.4; DB 7; Length 1641;  
 Best Local Similarity 84.7%; Pred. No. 3.2e-212;  
 Matches 1349; Conservative 0; Mismatches 181; Indels 62; Gaps 16;  
 QY 2 CCCACGCCCAAGCTCGCCAAACGGCGACACCATATCCGGTCTCAACGCCATATCAACGAG 61  
 DB 43 CCCACTGCCAGCTCGCCAAACGGCGACACCATATCCGGTCTCAACGCCATATCAACGAG 102  
 QY 62 GGGTCTCGGATTCCTTTGGCAGCGCGGTGGCAACCTCGGTTCAAGACCCCT 121  
 DB 103 GGGTCTCGGATTCCTTTGGCAGCGCGGTGGCAACCTCGGTTCAAGACCCCT 162  
 QY 122 GNGCGGTACTCTGGCTCGCTCAACGGCCAGAAAGT-----TACTTAGCGCCG---TGCATG 174  
 DB 163 GTGCCGTACTCCGGCTCGCTCGATGGCCAGAGTTCACTTCTACGGCCGCTTTCATG 222  
 QY 175 CAGCAGAACCCCGAGGGCAGCTTTGAAGAGAACCTTTGGCAAGACGGCACTCGACTTTGGTG 234  
 DB 223 CAGCAGAACCCCGAGGGCAGCTTACGAGGAGAACTCCCCAAGGCGAGCGCTCGACTTTGGTG 282

QY 235 ATCAGTCCAAAGGTGTTCCAGGCGGTGCTTCCAGAGTAGAGACTGCCTCAACATCAAC 294  
 DB |||||  
 DB 283 ATCAGTCCAAAGGTGTTGAGGCGGTGTCTCCGTCTAGCGAGGACTGTCTCACCATCAAC 342  
 QY |||||  
 DB 295 GTGGTGGCGCGCGCGGACCAAGGGGGGGGCAACCTCCCGGTCTAGCTCTGGATCTTT 354  
 DB |||||  
 DB 343 GTGGTGGCGCGCGCGGACCAAGGGGGGTGCAACCTCCCGGTGATGCTCTGGATCTTT 402  
 QY |||||  
 DB 355 GCGGTGGGTTCAGATCGGACGCCACCATCTTCCCTCCCGCCAGATGGTCAACCAAG 414  
 DB |||||  
 DB 403 GCGGCGGGTTCAGGTGGGTGGCAACAGCACCTTCCCTCCCGCCAGATGATCAACCAAG 462  
 QY |||||  
 DB 415 AGTGTGCTCATGGCAAG-CCATCATCCAGTGGCGGTCAACTACCGTGTGCTCGTGG 473  
 DB |||||  
 DB 463 AGCATTCGCATGGCAAGCCCATCATCCAGTGGCGGTCAACTACCGGTGTGCTCGTGG 522  
 QY |||||  
 DB 474 GGGTTCGTGGTGTGATGATCATCAAGCCGAGGCGAGCGGACCGCGGTTCGAAGGAC 533  
 DB |||||  
 DB 523 GGGTTCGTGGTGTGCGACGAGATCAAGGCGAGGCGAGTGCACACGCGCGTTCGAAGGAC 582  
 QY |||||  
 DB 534 CAGCGTTTGGGCATGACGTGGGTGGCAGACAAATTCGCGGTTCGGCGGCGACCCGAGC 593  
 DB |||||  
 DB 583 CAGCGCATGGCATGCGAGTGGGTGGCGGACAAATTCGCGGTTCGGCGGCGACCCGACC 642  
 QY |||||  
 DB 594 AAGGTGA-CATCTTTGGCGAG---GCGGGCAGCATGTCCGTGTGTTGGCACCTCATCTGG 649  
 DB |||||  
 DB 643 AAGGTGACCATCTTTGGCGAGTCTGCGGCGAGCATGTCCGTCTATGTGCCACATCTCTGG 702  
 QY |||||  
 DB 650 AACGACGGGACAAACATGACGATCAAGGGCAAGCGGTGTTCCGCGGCGCATCATCGAG--- 706  
 DB |||||  
 DB 703 AACGACGGGACAAACATGACGATCAAGGGCAAGCGGTCTTCCGCGGCGCATCATCGAGTCT 762  
 QY |||||  
 DB 707 GGAGCCATGTT---GCGGACCCCGTGGACGGCAGTACGGCAACGAGATCTACGACCTC 763  
 DB |||||  
 DB 763 GGGGCCATGTTACCGTGGACGCGGTGGACGGCGTCTACGGCAACGAGATCTTTGACCTC 822  
 QY |||||  
 DB 764 TTTGCTCGAGTGTGGTGTGGCAGCGCCAGCGCATCAAGTCTGGGTGCTTGGCGAGTGGC 823  
 DB |||||  
 DB 823 TTGGTCCGAGCGCGCGGTGGCAGCGCCAGCGCATCAAGTCTGGGTGCTTGGCGGTGTG 882  
 QY |||||  
 DB 824 ---AGCGACATCTGCTCGATGCCACCAACAACTCTCGGTGTTCTTCCGGTACTCTCTCG 880  
 DB |||||  
 DB 883 TCTAGCGACATCTTGGAGCGCGCCACCAACACCCCTGGGTCTTGGCGTACTCTCTCG 942  
 QY |||||  
 DB 881 TTGGCGTGTG-----ACTCCGCGCCGACGCGCAAGAACATCACCGATGATGATPACAG 934  
 DB |||||  
 DB 943 TTGGCGTGTGTTTATCTCTCCCGCGCCGACGCGGTGAACATCACCGACGATGTTGCC 1002  
 QY |||||  
 DB 935 TTGGTGGCGGACGGCAAGTATGCAAGCGTTCCCGTGTGATCTTGGCGACCGAGAACGAG 994  
 DB |||||  
 DB 1003 TTGGTCCGAGGGGCAAGTATGCAAGCGTTCTGTGTATCTGCGCGCGACAGAACGACGAG 1062  
 QY |||||  
 DB 995 GGCACCATCTTTGGCTC-----TTGAACTGACCAAGAACTCTCAGCGCCGTCTTAC 1047  
 DB |||||  
 DB 1063 GGCACCTCTTTGGCACCTCTTCTTTGAACTGACCAAGATGCGGAGCGCCGCGAGTAC 1122  
 QY |||||  
 DB 1048 TTCAAGCAG---TTCATCCAGCCAGCGACGCGAGATCGACACCTTGTATGCGCGGTAC 1104  
 DB |||||  
 DB 1123 TTACGCGAGTCTTTGTTCACGCGCAGCGCGGAGTTCGACACGTGATGATGCGCGTAC 1182  
 QY |||||  
 DB 1105 CCCGAGGACATCACCCAGGG---TCCGTTGGACACGGTT-----CAGCGTCAACCGC 1154  
 DB |||||  
 DB 1183 CCCGAGGACATCACCCAGGGTTCTCCGTTTCGACACCGGGTGTCTCAACGCCCTCACCCG 1242  
 QY |||||  
 DB 1155 CAGTTCAAGAGNATC---GCGGTGCTCGCGACCTTGATTCATTCACGCCCGCGGTAC 1211  
 DB |||||  
 DB 1243 CAGTTCAAGAGNATCTCTCGGTGCTCGCGGACCTTGACCTTCATCCAGCCCGCTCGCTAC 1302  
 QY |||||  
 DB 1212 TTCTCTCAACCACTTCAGGGCGGACCAAGTACTCTGTTCTC-----AAGCAGTCTGGG 1265  
 DB |||||  
 DB 1303 TTCTCTCAACCACTACACGGCGGACCAAGTACTCTCTTCTTAAGCAGGTCTCTGGC 1362

QY 1266 TTGCAATCATGGGCACTTCCATGCGCAAGACATTTGTGGCAGGACTACTTGTGGGA 1325  
 DB 1363 TTGCGGTGTCTCGGAAGTTCCTCACTCCAAAGACATTTGTTCAGGACTACTTGTGGGC 1422  
 QY 1326 AGCGGAGCGTCACTACAAACGCGTTTATCGGTTCCGACCGACTTGGACCCCAAC 1395  
 DB 1423 AGCGGCTCGCTCATCTACAAACGCGTTTATCGGTTTGCACGGACTTGGACCCCAAC 1482  
 QY 1386 ACCGCGGGTGTGTGGTGAATCGTCCCAAGTACACACGAGCAGC-----CAGGGCAACAAC 1439  
 DB 1483 ACCGCGGGTGTGTGGTGAATGSCCGAGTACACACGAGCTCTCAGTCTGGCAACAAC 1542  
 QY 1440 TTGATGATGATCAACGCTTGGGTTGTATACCGGGCAAGACAACTTCGACCGCTGGC 1499  
 DB 1543 TTGATGATGATCAACGCTTGGGTTGTATACCGGGCAAGACAACTTCGACCGCGGC 1602  
 QY 1500 TACGACGCGTTGATGACCAACCGTCTTGT 1531  
 DB 1603 TACGACGCGTTGTCTCAACCGCGCTCTTT 1634

RESULT 4  
 AAX33111  
 ID AAX33111 standard; DNA; 1649 BP.  
 XX AAX33111;  
 XX  
 DT 23-JUN-1999 (first entry)  
 XX  
 DE Candida rugosa lipase 1 gene.  
 XX  
 KW Candida rugosa; lipase 1; LIPI; industrial bioconversion; ss.  
 XX  
 OS Candida rugosa.  
 XX  
 PN WO9914338-A1.  
 XX  
 XX 25-MAR-1999.  
 XX  
 PF 16-SEP-1997; 97WO-NL000524.  
 XX  
 PR 16-SEP-1997; 97WO-NL000524.  
 XX  
 XX (UNIL ) UNILEVER NV.  
 XX  
 PI Brocca S, Schmidt-Dannert C, Lotti M, Alberghina L, Schmid R;  
 XX WPI; 1999-229539/19.  
 DR  
 XX Synthesis and functional overexpression of a Candida rugosa lipase gene  
 PT coding for a major industrial lipase.  
 XX  
 PS Claim 13; Page 23-28; 44pp; English.  
 XX  
 CC The present sequence represents the native Candida rugosa lipase 1 gene.  
 CC Lipases produced by Candida rugosa are extensively used in industrial  
 CC bioconversions, and the pure lipase 1 can be used in a process requiring  
 CC high specificity toward acyl chains shorter than 14C. Lipase 1, free of 2  
 CC -5, can be obtained without using extensive and expensive working up  
 CC procedures. Pure lipase 1 exhibits higher activity toward capriate than  
 CC toward palmitate  
 XX  
 SQ Sequence 1649 BP; 317 A; 536 C; 493 G; 301 T; 0 U; 2 Other;

Query Match 64.7%; Score 991.6; DB 2; Length 1649;  
 Best Local Similarity 84.2%; Pred. No. 2.2e-206;  
 Matches 1336; Conservative 0; Mismatches 186; Indels 64; Gaps 17;

QY 2 CCCACGCCCAAGCTCGCCAAACGGGACACCATCACCGGTCTCAACGCCATCATCAACGAG 61  
 DB 49 CCCACGCCCAAGCTCGCCAAACGGGACACCATCACCGGTCTCAACGCCATCATCAACGAG 108  
 QY 62 GCGTTCTCGGCATTCCTTTGCGGAGCGCGGNGGGCAACCTCGCTTCAAGGACCCT 121

DB 109 GCGTCTCTCGGCATTCCTTTTGGCGAGCGCGGTGGCAACCTCCGCTTCAAGGACCCC 168  
 QY 122 GTGCCGTACTCTGGCTCGCTCAACGGCCAGAAGTT-----ACTTACGGCCCG--TGCATG 174  
 DB 169 GTGCCGTACTCTGGCTCGCTCGATGGCCAGAAGTTTACAGCTACAGGCCCGAGCTGATG 228  
 QY 175 CAGCAGAAACCCGAGGCGCAGTTTGAAGAGAACCTTGGCAAGACGGCAGCTCGACTTGGTG 234  
 DB 229 CAGCAGAAACCCGAGGCGACCTTACGAGAGAACCTCCCAAGGAGCGCTGACTTGGTG 288  
 QY 235 ATGCACTCAAGGTGTTCAGGCGGTCTTCCCAGAG--TCAGGACTCGCTTCAACATCAA 293  
 DB 289 ATGCACTCAAGGTGTTCAGGCGGTGAGCCCGGAGCAGCAGGAGCTGTCTCACATCAA 348  
 QY 294 CGTGGTGGCGCGCGGGGACCAAGGGCGGCGCAACTCCCGGTCAATGTCTGGATCTT 353  
 DB 349 CGTGGTGGCGCGCGGGGACCAAGGGCGGTGCAACTCCCGGTGATGTCTGGATCTT 408  
 QY 354 TGGCGGTGGGTTCAGATCGGAGCGCCACCATCTTCCCTCCCGCCAGATGTCACCAA 413  
 DB 409 TGGCGCGGTTCAGATGAGTGGTGGCAGCAGCTTCCCTCCCGCCAGATGTCACCAA 468  
 QY 414 GAGTGTCTCATGGGCAAG--CCATCATCCAGTGGCGGTCAACTTACCGTGTTCGCTCGTG 472  
 DB 469 GAGCATTTGCGATGGCAAGCCCATTCACAGTGAGCGTCAACTACCGGTGTCTGCTG 528  
 QY 473 GGGGTTCCTGGCTGTGATGACATCAAGCGCGAGGCGAGCGGGAACCGCGGCTTGAAGGA 532  
 DB 529 GGGGTTCCTGGCTGGCGACGAGATCAAGCGCGGCGAGTGCACCAACCGCGCTTGAAGGA 588  
 QY 533 CCAGCGTTTGGGCGATCGAGTGGGTGCGAGCAACATTCGCGGTTCGGCGCGGACCGGAG 592  
 DB 589 CCAGCGTTTGGGTATGAGTGGGTGGGCAACATTCGCGGTTCGGCGCGGACCGGAC 648  
 QY 593 CAAAGTGCAC--ATCTTTGGCGAG---GCGGCGAGCATGTCCGTGTTCGACACCTCATCTG 648  
 DB 649 CAAAGTGCAGATCTTTGGCGAGAGCGGCGAGCATGTCCGTGTTCGACATTCGCTG 708  
 QY 649 GAACGAGCGGGAACAACACGATCAAGGGCAAGCGGTGTCGCGCGGGGCAATGAG-- 706  
 DB 709 GAACGAGCGGGAACAACACGATCAAGGGCAAGCGGTGTCGCGCGGGCATCATGAGAG 768  
 QY 707 -GGAGCATGTGGCGCG---GACCCGTGAGCGGCGAGTGCAGCAACGAGATCTACGACCT 762  
 DB 769 CGGGCGATGTGTCGAGCGAGCGCGGTGAGCGCATCTACGGCAACGAGATTTTACCT 828  
 QY 763 CTTTGTCTCGAGTGTGCTGTGGCAGCGCCAGCGCAAGCTCGCGTGTTCGCGAGTGC 822  
 DB 829 CTTTGGCGTCCGAACGCGGGCTGCGGAGCGCCAGCGAAGCTCGCGTGTTCGCGGTGT 888  
 QY 823 G---AGCGACACTTGTCTGATGCGACCAACAACTCTCTGGTTCCTGGCGTACTCCTC 879  
 DB 889 GAGCAGCGACAGTGTGGAGGAGCGCCACCAACAAACACCCCTGGGTTCCTGGCGTACTCCTC 948  
 QY 880 GTTTCGGTTGTACT-----CCGCGCCCGAGCGAGACATCACCGATGACATGACAA 933  
 DB 949 GTTTCGGTTGAGCTACTCTCCCGCGCGCGCGGTGAACATCACCGAGCATGTACGC 1008  
 QY 934 GTTTCGCGCGAGCGGCAAGTATGCAAGCGTTCCTCGTGTATTCATTTGGCGACCAAGACGA 993  
 DB 1009 CTTTGTGGCGAGGCGAGTATGCCAACATCCTCTGTATCATCGGCGACCAAGACGA 1068  
 QY 994 GGGCACCATCTTTGGC-----TCTTGAACGTGACCAAGATGTTCAGCCCGGTCTTA 1046  
 DB 1069 GGGCACCTCTTTTGGGACCCAGCAGCTTGAACTGACCAAGATGCGGCGCGGAGTA 1128  
 QY 1047 CTTCAAGCAG---TTTATCCAGCGCAGCGCGGAGTGCACCTTTCATGGCGGCTA 1103  
 DB 1129 TTTCAAGCAGAGCTTTGTCCACGCGGAGCGGAGATCGACACGTTGATACGCGCTA 1188  
 QY 1104 CCCCAGGACATCAACCCAGG---GTCCGTTTCACACGCTT-----CAACGCTCACCCC 1153

Db 1189 CCCGGCGCATCACCCAGGCGACCGCGTTTCGACACGGGTATTCTCAACGGCTTCAACCC 1248  
 Qy 1154 GCAGTTCAAGAGAT---CGCGTGTCTCGGCGACCTTGATTCATCCAGCGCCGCGCTA 1210  
 Db 1249 GCAGTTCAAGAGATCAGCGCGGTCTCGGCGACCTTTGCTTACGCTTGTCTCGCTA 1308  
 Qy 1211 CTTCCTCAACACACTTCCAGGCGGCGCACCAAGTACTCGTTCTCA-----AGCAGCTCGG 1254  
 Db 1309 CTTCCTCAACACACTTCCAGGCGGCGCACCAAGTACTCGTTCTCAAGCAGCTCAGCGG 1368  
 Qy 1265 GTTGCCCAATCATGGGCGCTTCCATGCCAACAAGCATTTGTGGCAGACTACTTTGTTGGG 1324  
 Db 1369 CTTCGCGGTCTCGGAAGGTTCCACTCAACGACATTTCTTCAGGAGTACTTTGTTGGG 1428  
 Qy 1325 AAGCGGAGCGGTCTATCAACAACGCGTTTATCGGTTCCGACCGACTTGGACCCCAA 1384  
 Db 1429 CAGCGGCTCGCTCATCTACAAACGCGTTTCAATGCGTTTCCACGAGCTTTGGACCCCAA 1488  
 Qy 1385 CACCGCGGGTGTGTGTGAACCTGGCCCAAGTACACACGACGAGC-----CAGGGCAACA 1437  
 Db 1489 CACCGCGGGTGTGTGTGAAGTGGCGCGAGTACACACGACGAGCGCAGCNGGCAACA 1548  
 Qy 1438 ACTTGATGATCAACGCGCTTGGGCTTGTATACCGGCAAGGACAACTTCCGACCGCTG 1497  
 Db 1549 ACTTGATGATCAACGCGCTTGGGCTTGTATACCGGCAAGGACAACTTCCGACCGCGG 1608  
 Qy 1498 GCTACGAGCGGTTGATGACCAACCG 1523  
 Db 1609 GCTACGAGCGGTTGTTCTCAACCG 1634

RESULT 5

ABX95908  
 ID ABX95908 standard; DNA; 1641 BP.

AC AC

ABX95908;

15-JUL-2003 (first entry)

Candida rugosa lipase 5 DNA.

Lipase 5; gene; ds; mutant.

Candida rugosa.

Key Location/Qualifiers  
 CDS 1..1641  
 /tag= a  
 /product= "Lipase 5"  
 /partial  
 /note= "No start or stop codon shown"

EP1288294-A2.

05-MAR-2003.

26-APR-2002; 2002EP-00009616.

31-AUG-2001; 2001US-00943857.

(SINI-) ACAD SINICA.

Tang S, Lee G, Shaw J;

WPI; 2003-395476/38.

P-P8DB; ABU09073.

Isolated mutant nucleic acid encoding Candida rugosa lipase, useful for the preparation of Candida rugosa lipase for biocatalytic applications.

Claim 25; Page 8-9; 33pp; English.

The invention relates to an isolated mutant nucleic acid encoding a

CC Candida rugosa lipase polypeptide. The DNA has a sequence having at least  
 CC 80% identity to a wild-type DNA encoding Candida rugosa lipase. The  
 CC invention also relates to a microorganism comprising the DNA, where the  
 CC microorganism is a bacterium or yeast, preparing a mutant DNA encoding a  
 CC C. rugosa lipase and a chimeric C. rugosa lipase comprising a substrate  
 CC interacting domain of a first C. rugosa lipase and a non-substrate  
 CC interacting domain of a second C. rugosa lipase. The method is useful for  
 CC preparing a mutant DNA encoding a Candida rugosa lipase. The DNA is  
 CC useful in the large scale manufacture of Candida rugosa lipase which is  
 CC useful for biocatalytic applications. This sequence represents DNA  
 CC encoding Candida rugosa lipase 5  
 XX  
 SQ Sequence 1641 BP; 302 A; 543 C; 444 G; 352 T; 0 U; 0 Other;

Query Match 63.2%; Score 968.6; DB 7; Length 1641;  
 Best Local Similarity 82.6%; Pred. No. 2.3e-201;  
 Matches 1309; Conservative 0; Mismatches 214; Indels 62; Gaps 15;

Qy 2 CCCACCGCCAAAGCTCGCCAAACGGCGACACCATCAACGGTCTCAACGGCATCATCAACGAG 61  
 Db 43 CCCACTGCCAGCTCGCCAAACGGCGACACCATCAACGGTCTCAACGGCATCATCAACGAG 102  
 Qy 62 GCGTTCTCTCGCATTTCCCTTTGCCAGCGCGGTGGGCAACCTCCGCTTCAAGGACCCCT 121  
 Db 103 GCGTTCTCTCGCATTTCCCTTTGCCAGCGCGGTGGGCAACCTCCGCTTCAAGGACCCCT 162  
 Qy 122 GTGCCGTACTCTGGCTCGCTCAACGGCGCAGAAAGTTACTTACGGCC-----CGTGCATG 174  
 Db 163 GTGCCGTACTCTGGCTCGCTCAACGGTCAATCTTCAACGGCTACGGTCCGCTCTTTCGATG 222  
 Qy 175 CAGCAGAACCCCGAGGGGACGTTTGAAGAGAACCTTGGCAGAGCGGCACTCGACTGTGTG 234  
 Db 223 CAGCAGAACCCCGAGGGGACCTTACGAGGAGAACCTTCCCAAGGTGGCGCTTACTTGTGTG 282  
 Qy 235 ATGCAGTCCAAAGGTGTTCCAGCGGTGTTCCCAAGTGAAGGACTGCTCACCATCAAC 294  
 Db 283 ATGCAGTCCAAAGGTGTTCCAGCGGTGTTTCCCAAGTGAAGGACTGCTCACCATCAAC 342  
 Qy 295 GTGGTGGCGCGCGCGGACCAAGCGGGCGCCAACTCCCGGTGATGCTCTGTGATCTTT 354  
 Db 343 GTGGTGGCGCGCGCGGACCAAGCGGGCGCCAACTCCCGGTGATGCTCTGTGATCTTT 402  
 Qy 355 GCGGTGGGTGATCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 414  
 Db 403 GCGGTGGGTGATCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 462  
 Qy 415 AGTGTGCTCATGGGCAAG--CATCATCCAGTGGCGGTCAACTACCGTGTGCTCTGTGG 473  
 Db 463 AGTGTGCTCATGGGCGAGCCCATCATCCAGTGGCGGTCAACTACCGTGTGCTCTGTGG 522  
 Qy 474 GGGTTCTTGGTGTGATGACATCAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 533  
 Db 523 GGTTCCTTGGCGGTCCGGAATCAAGGCGGAGGCGAGTCCAAATGCGCGGCGTCAAGGAC 582  
 Qy 534 CAGCGTTTGGGCGATCGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 593  
 Db 583 CAGCGCTTGGGCGATCGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 642  
 Qy 594 AAGGTGA-CATCTTTGGCGAG---GCGGCGAGCATGTCCGTTGTTGGCCACCTCATCTGG 649  
 Db 643 AAGGTGACCACTTTTGGCGAGTCTCGGGGCGAGCATGTCCGTTGTTGGCCACCTCATCTGG 702  
 Qy 650 AACGACGGCGCAACAACGTTAAGGGGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 706  
 Db 703 AATGGCGCGCAACAACGTTAAGGGGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 762  
 Qy 707 GGAGCCATGGTGGCG---GACCCGTTGACGCGACGATACGCGCAAGAGATCTACGACCTTC 763  
 Db 763 GGAGCCATGGTGGCGTCTGACCCGTTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 822  
 Qy 764 TTTGTCTCGAGTGTGGTGTGTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 820  
 Db 823 TTGGTGGCTTCTACGGGCTGCAGCAGTGCACGCAACAAGCTTGGCTGTGGTGTGGTGTCTT 882







Db 163 GTCCGCTACTCGCGTCTGCTCAACGGCCAGAGTTTACCTCTTACGGCCGCTCTTGATG 222  
 Qy 175 CAGCAGAACCCGAGGCGCAGTTTGAAGAGAACCTTGGCAAGACGGCACTCGACTTGTGTG 234  
 Db 223 CAGATGAACCTATAGGCTCGTTTGAAGACACACTTCCCAAGATGCGCTTGAATGGTG 282  
 Qy 235 ATGAGTCCAAAGTGTTCAGGCGGTGCTTCCCAAGAGTGAAGTGCCTCACCATCAAC 294  
 Db 283 CTCAGTCCAAAGTCTTCCAAAGTGGTCTTCCCAAGACGAGGAGTGTCTCACCATCAAC 342  
 Qy 295 GTGCTGGCGCGCGGCGGACCAAGCGGCGGCCAACCTCCGGTCACTGCTCTGATCTTT 354  
 Db 343 GTGATCCGGCGCGCGGCGGACCAAGGCGGAGTGTGCTCTCCGGTGAATGCTCTGATCTTT 402  
 Qy 355 GCGGTGGGTGTGAGATCGGCGGCGGCCCAACCATCTTCCCTCCCGCCAGATGCTCAACAG 414  
 Db 403 GCGGTGGGTGTGAGCTTGGCGGCTCCAGCTCTTTCCAGGAGACAGATGGTGGCCAG 462  
 Qy 415 AGTGTGCTCANTGGGCAAGCC-ATCATCCAGTGGCGGTCAACTACCGTGTGCTCTGCTGG 473  
 Db 463 AGCGTGTCTCATGGGTAAACCGGTGATCCACGTGAGCATGAATACGCGTGGCGTCAATGG 522  
 Qy 474 GGGTCTCTGGCTGTGATGATCAATCAAGCCGAGCGGAGCGGGAACCGCGCTTGAAGGAC 533  
 Db 523 GGGTCTCTGGCGCGCGCGGATCCAGAACGAGGAGCGGGAACCGCGCTTGCATGAC 582  
 Qy 534 CAGCGTTTGGCATGCACTGGGTGGGAGACAAATTCGCGGTTCGGCGCGGACCGGAGC 593  
 Db 583 CAGCGTTTGGCCATGCACTGGGTGGGAGACAAATTCGCGGTTCGGCGCGGACCGGAGC 642  
 Qy 594 AAGTGA-CATCTTTGGGAG---GCGGGAGCATGTCCGTGTGTGCCACCTCATCTGG 649  
 Db 643 AAGTGACCATATACCGCGAGTCTCGGGCAGCATGTTCGACGTTTGTGACCTTGTGTGG 702  
 Qy 650 AAGCGGGGCAACACGCTACAGGCGAGCGGTGTTTCCGCGCGGACATCATGAGGGA 709  
 Db 703 AAGCGGGGCAACACGCTACAGGCGAGCGGTGTTTCCGCGCGGACATCATGAGGAGT 762  
 Qy 710 GCC-----ATGCTGCGGACCGGTGGAGCGGACGCTAGCGCAACAGATCTACGACCTC 763  
 Db 763 GCGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 822  
 Qy 764 TTTGCTCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 820  
 Db 823 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 882  
 Qy 821 GCGAGCGACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 880  
 Db 883 TCTCAGGACACGTTGTACAGGCGGACGAGCAGCGCGCGGCTGCTGCTGCTGCTGCTGCTG 942  
 Qy 881 TTGCGGTTGT-----ACTCCGCGCGCGGACGCAAGAACATCAGCGATCAGATGTACAG 934  
 Db 943 TTGCGGTTGTCTATCTCCCGCGCGCGGACGCAAGAACATCAGCGATCAGATGTATGCC 1002  
 Qy 935 TTGCTGCGGCGGCAAGATGATCAAGCGTTCGCGTATCATTTGGGACCGAGAACGACGAG 994  
 Db 1003 TTGCTGCGGCGGCAAGATGATCAAGCGTTCGCGTATCATTTGGGACCGAGAACGAG 1062  
 Qy 995 GGCACATCTTTGG-----CTCTTGAACGTGACCAAGATGCTCAGGCGCGGTGTAC 1047  
 Db 1063 GGCATTTGTTTGGGCTCTCTTCTTGAACGTGACCAAGATGCTCAGGCGCGGTGTAC 1122  
 Qy 1048 TTCAAGCAG---TTCAATCCAGCGGACGCGGAGATGACACCTTGTATGCGCGGTAC 1104  
 Db 1123 TTCAAGCAGCTCTTATCCAGCGGAGGATGCGGAGATGACACCTTGTATGCGCGGTAC 1182  
 Qy 1105 CCCCAGGACATCACCAGG---TCCGTTGACACGG-----TTCAACGCTCACCCTG 1154  
 Db 1183 ACCAGGACATCACCAGGTTCTCCGTTGACACCGGATCTTCAATGCCATCACCCTG 1242  
 Qy 1155 CAGTTCAAGAGATC---GCGGTGCTCGGCGGACCTTGTGATTCATCAGCGCGCGGTAC 1211

Db 1243 CAGTTCAAACGAGTCTCTGCTTGGCGACCTTGGCTTACGCTTGGCGTGGCTAC 1302  
 Qy 1212 TTCTCAACACCTTCCAGGGGGGCGACCAAGTACTCGTTCTCTC-----AAGCAGCTGGG 1265  
 Db 1303 TTCTCAACCTTACACAGGGGGGCGACCAAGTACTCGTTCTCTCTAAGCAGCTTCTGGG 1362  
 Qy 1266 TTGCCAATCATGGGCGACCTTCCATGCCAACGACATGTGTGGCAGGAGTACTTGTGGGA 1325  
 Db 1363 TTGCCCGCTTGTGGGACCTTCCAGCGGACGACATCATTTGGCAGGAGTACTTGTGGG 1422  
 Qy 1326 AGCGGAGCTCATCTATCAACAAACGCGTTTATCGCTTCCGACCGACTTGGACCCCAAC 1385  
 Db 1423 AGCGGAGTGTGATCTACAAACGCGTTTCAATTGGCTTTCGCAACGAGCTCGACCCGAC 1482  
 Qy 1386 ACCGCGGGTGTGTGTGAATCTGGCCCAAGTACACGAGAGC-----CAGGGGCAAC 1439  
 Db 1483 AAGCGGGCTTGTGGACCAACTGGCCCAAGTACACGAGAGCTCTCAGTCTGGCAAC 1542  
 Qy 1440 TTGATGATGATCAACGCTTGGGCTTGTACACCGGCAAGGACAACTTCCGACCGCTGGC 1499  
 Db 1543 TTGATGAGATCAACGCTTGGGCTTGTACACCGGCAAGGACAACTTCCGCGGATGCG 1602  
 Qy 1500 TAGCAGCGTGTGATGACCAACCGCTTCTTTGT 1531  
 Db 1603 TAGCGCGCTCTTTTCCAACCGCGCTCTTT 1634

## RESULT 8

AAK33112

ID AAK33112 standard; DNA; 1647 BP.

XX AAK33112;

XX 23-JUN-1999 (first entry)

XX Synthetic lipase 1 gene.

XX Candida rugosa; lipase 1; Lip1; industrial bioconversion; ss.

XX Synthetic.

XX Candida rugosa.

XX WO9914338-A1.

XX 25-MAR-1999.

XX 16-SEP-1997; 97WO-NL000524.

XX 16-SEP-1997; 97WO-NL000524.

XX (UNIL ) UNILEVER NV.

XX Brocca S, Schmidt-Dannert C, Lotti M, Alberghina L, Schmid R;

XX WPI; 1999-229539/19.

XX Synthesis and functional overexpression of a Candida rugosa lipase gene coding for a major industrial lipase.

XX Disclosure; Page 23-28; 44pp; English.

CC The present sequence represents a synthetic lipase 1 gene derived from the native Candida rugosa lipase 1 gene. Lipases produced by Candida rugosa are extensively used in industrial bioconversions, and the pure lipase 1 can be used in a process requiring high specificity toward acyl chains shorter than 14C. Lipase 1, free of 2-5, can be obtained without using extensive and expensive working up procedures. Pure lipase 1 exhibits higher activity toward caprinic than toward palmitic

XX Sequence 1647 BP; 377 A; 400 C; 364 G; 506 T; 0 U; 0 Other;

Query Match

Best Local Similarity 42.5%; Score 651.8; DB 2; Length 1647;

Pred. No. 2.4e-132;

Matches 1104; Conservative 0; Mismatches 417; Indels 62; Gaps 14;

QY 2 CCACCGCCAAAGCTCGCAACGCGGACACCATCACCGGTCTCAACGCCATCATCAACGAG 61  
 Db |||||  
 QY 49 CCAACGCCCACTTTGGCTAAACGGTGACACCATCACCGGTTTGAACGCCATCATCAACGAA 108  
 Db |||||  
 QY 62 GCGTTCCTCGGCATTCCTTTGCGAGCCGCGGTGGCAACCTCCGCTTCAAGGACCT 121  
 Db |||||  
 QY 109 GCCTTCTTGGGTATTCATTTGCGGAACCAACAGTTGGTAACCTTGAGATTAAGGACCCA 168  
 Db |||||  
 QY 122 GTGCGGTACTCTGCTCGCTCAACGCGCAGAGTTACTTACGGCC-----CGTGCATG 174  
 Db |||||  
 QY 169 GTTCCATATCTCGGTTCTTGATGGTCAAAAGTTCACTTCTACGGTCCATCTGTATG 228  
 Db |||||  
 QY 175 CAGCAGAACCCCGAGGACAGCTTTGAAGAGAACCTTGGCAAGCGGACCTCGACTTGGTG 234  
 Db |||||  
 QY 229 CAACAAACCCAGAAGGTACTCAAGAGAAACTTGGCAAGGAGCTTTAGATCTGTGTT 288  
 Db |||||  
 QY 235 ATGCAATCCAAAGTTTTCGAAGCTGTTTCTCCATCTCTGAACTGTTGACCAATTAAT 348  
 Db |||||  
 QY 295 GTGCTGCGCGCGCGGACCAAGCGCGCGCCCAACCTCCCGGTATGCTCTGATCTTT 354  
 Db |||||  
 QY 349 GTTGTAGACCAACCGGACAAAGGCTGTGCCAACTTGGCCAGTTATGTTGTGATCTTT 408  
 Db |||||  
 QY 355 GCGCGTGGGTTTGAATCGGAGCCCAACATCTTCCCTCCCGCCAGATGTCACCAAG 414  
 Db |||||  
 QY 409 GGTGGTGGTTTGAAGTTGGTGTACTAGTACTTCCCTCCAGCCCAATGATTACCAAG 468  
 Db |||||  
 QY 415 AGTGTCTCATGGGCAAGCC-ATCATCCAGTGGCGGTCAACTACCGTGTTCCTCGTGG 473  
 Db |||||  
 QY 469 TCTATTGTATGGGTAAGCCCAATCATCAACGTTTCTGTCACATACAGATCTCGAGCTGG 528  
 Db |||||  
 QY 474 GGTGTTCTTGGTGTGTATGATCACTCAAGCGCGAGGCGAGCGGAAACCGCGGTGTAAGAAC 533  
 Db |||||  
 QY 529 GGTGTTCTTGGTGTGTATGATCACTCAAGCGCGAGGTTCTGCCAAGCGGTTTGAAGGAC 588  
 Db |||||  
 QY 534 CAGCGTTTGGGATGAGTGGTGGGAGCAACATTTGCGGGTTGCGGCGGACCGGAGC 593  
 Db |||||  
 QY 589 CAAAGATTGGGTATGCAATGGGTGGTGTGACCAATTTGCTGTCTTGGTGGTATCCAAT 648  
 Db |||||  
 QY 594 AAGGTGAC-ATCTTTGGCGA---GGCGGACAGTGTGCTGTGTGCGCACCTCATCTGG 649  
 Db |||||  
 QY 649 AAGGTACTATCTTTGGTGAATCTGCTGTTTCTATGTCGCTCATGTGTGACATTTGGG 708  
 Db |||||  
 QY 650 AACGACGCGGCAACACACGTACAAAGGCAAGCGGTTTCCGCGCGGCAATCAGCA--G 706  
 Db |||||  
 QY 709 AACGACGCGTGAACACACTTCAAGGTAAGCCATTGTTTCAAGAGCTGTTATCATGCAATCT 768  
 Db |||||  
 QY 707 GGAGCATGGTGC---CGGACCGGTGGAGCGGACGTCAGGCAACGAGATCTACGACCTC 763  
 Db |||||  
 QY 769 GGTGCTATGGTTCATCTGACGCGGTGCGAGCGGTATCTACGGTAACGAAATTTTGACTTG 828  
 Db |||||  
 QY 764 TTGTCTCGAGTCTGCTGTGCGAGCGGACGCAAGCTGCGGTGTTGGCGAGTG-- 821  
 Db |||||  
 QY 829 TTGGCTTCAACCGCTGGTGGTGGTCTGCTCTGACAAAGTGGCTGTTTGGAGAGTGT 888  
 Db |||||  
 QY 822 -CGAGCGACACCTTGTCTGATGCCAACAAACACTCTCTGGGTCTTGGGCTACTCTCG 880  
 Db |||||  
 QY 889 TCTTCTGACACTTTGGAAGAGCGCCCAACAAACACCCCTGTTCTTGGTGTACTCTCTC 948  
 Db |||||  
 QY 881 TTGCGGTTGTACT-----CCGCGCGGAGCGGCAAGAAATACCATCATGATGATCAAG 934  
 Db |||||  
 QY 949 TTAAGATTGTCTTACTTTGCGAAGCAGCGGTGTTAATCATCCGACGATGTAGCT 1008  
 Db |||||  
 QY 935 TTGGTGGCGACGCGAGTATGCAAGGTTCCCGTGTATCTTGGCGACCAAGACGAG 994  
 Db |||||  
 QY 1009 TTGGTTAGAAGTATGATGCAACATCCCTGTTATCATCGTGCACAAAGAGGAA 1068  
 Db |||||  
 QY 995 GGCACCATCTTTGGCTC-----TTGAACGTGACCAAGAAATGCTCAGGCCCGGTCTAC 1047  
 Db |||||  
 QY 1069 GGTACCTTCTTTGGTACTTCTTCTTGAAGGTTTACCACTGATGTCACCAAGCCAGAGATAT 1128  
 Db |||||

RESULT 9

AAH43625

ID AAH43625 standard; cDNA; 1950 BP.

XX

AC AAH43625;

XX

DT 21-JAN-2002 (first entry)

XX

DE Synthetic lipi gene.

XX

Lipase; lipi; variant; ripening form; serine; 16-18C acyl chain; ss.

XX

Candida rugosa.

OS

Key Location/Qualifiers

FT CDS

175..1779

/product= "lipi"

/transl\_except= (pos:292..294,aa:Glu)

/transl\_except= (pos:301..303,aa:Gln)

/transl\_except= (pos:532..534,aa:Thr)

/transl\_except= (pos:553..555,aa:Thr)

/transl\_except= (pos:565..567,aa:Arg)

/transl\_except= (pos:802..807,aa:Thr Thr)

/transl\_except= (pos:1024..1026,aa:Glu)

/transl\_except= (pos:1114..1116,aa:Thr)

/transl\_except= (pos:1288..1290,aa:Thr)

/transl\_except= (pos:1678..1680,aa:Gly)

/transl\_except= (pos:1753..1755,aa:Phe)

FT

XX

PN

XX

PD

XX

05-SEP-2001.

XX

02-FEB-2001; 2001EP-00200375.

PF

EF1130100-A1.

XX 14-FEB-2000; 2000EP-00200513.  
 XX (UNIL ) UNILEVER NV.  
 PA (UNIL ) UNILEVER PLC.  
 XX Brocca S, Bornscheuer UT, Pleiss J, Schmid RD, Schmid U;  
 PI Schmitt J;  
 XX WPI; 2001-649825/75.  
 DR P-PSDB; AAB47622.  
 XX Modified lipolytic enzymes with altered substrate specificity, useful for  
 PT biocatalytic applications comprising high specificity towards carbon 16  
 PT and carbon 18 acyl chains.  
 XX Disclosure; Fig 1; 33pp; English.  
 XX This sequence encodes a parent lipase, lip1, derived from *C. rugosa*. The  
 CC lipase of the invention is a variant of this parent lipase, with altered  
 CC properties. The variant is the ripening form of *C. rugosa* lipase selected  
 CC from pre, pro, prepro or mature lipase, in which 60% or less of the CTG  
 CC codons encoding serine in the native *C. rugosa* sequence, are replaced by a  
 CC universal codon for serine. The modified nucleic acid sequence is further  
 CC modified, such that lipase variant exhibits an altered property. The  
 CC modified lipase is useful in a process requiring high specificity towards  
 CC 16-18C acyl chains  
 XX  
 SQ Sequence 1950 BP; 472 A; 478 C; 420 G; 580 T; 0 U; 0 Other;  
 Query Match 42.5%; Score 651.8; DB 4; Length 1950;  
 Best Local Similarity 69.7%; Pred. No. 2.5e-132;  
 Matches 1104; Conservative 0; Mismatches 417; Indels 62; Gaps 14;  
 2 CCCACGCCCAAGCTCGCCAAAGCGCACACCATCACGGTCTCAACGCCATCATCAAGCAG 61  
 178 CCACGCCCACTTTGGTAAAGGTGACACCATCACGGTCTCAACGCCATCATCAAGCAG 237  
 62 GCGTTCCTCGGCATTCCTCTTGGCGAGCCGCGTGGGCAACCTCCGCTTCAAGAGCCCT 121  
 238 GCCTTCTTGGGTATTCATTTGGCAACACACGATTTGGTAACTTGAGATTCAAGGACCCA 297  
 122 GTCCGCTACTCGTCTGCTCAACGGCCAGAGTTACTTACGGCC-----CGTGATG 174  
 298 GTTCCATACCTCCGGTTCCTTGGATGGTCAAAAGTTCACCTTCTTACGGTCCATCTTGATG 357  
 175 CAGCAGAACCCCGAGGCGCAGTTTGAAGAGAACTTGGCAAGCGCACTCGACTTGGTG 234  
 358 CAACAAACCCAGAGGTACTAGAGAGAACTTGGCAAGGCGAGCTTTAGATCTGGTT 417  
 235 ATGCAATCCAAAGTTTTCGAAGCTGTTCTCCATCTCTGAGACTGTTGACCATTAAT 477  
 295 GTGGTGGCGCGCGGCGCACCAAGGCGGCGCAACTCCCGGTCTAGCTCTGGATCTTT 354  
 478 GTTGTGTAGACCAACCGGCGCAAAAGGCTGGTGGCAACTTGGCAAGTATGTGTGATCTTT 537  
 355 GGGGTGGGTGTTGAGATCGGCAGCCCAACCATCTTCCCTCCCGCCAGATGGTCAACAAG 414  
 538 GGTGGTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 597  
 415 AGTGTGCTCATGGGCAAGCC-ATCATCCACGTCGGCGTCAACTACCGTGTGCTCGTGG 473  
 598 TCTATTGCTATGGTAAAGCAATCATCCAGTTTCTGTCACTACAGAGTCTCGAGCTGG 657  
 474 GGGTCTTGGTGGTGAATCAAGCGCGAGGCGAGGCAACCGCGCTTTGAAGGAC 533  
 658 GGTCTTGGTGGTGAATCAAGCGCGAGGCGAGGCTTCTGCCAAACCGCGTTTGAAGGAC 717  
 534 CAGCGTTTGGGCAATGAGTGGGTGGGAGACAACTTGGCGGGTTCGGCGCGAGCCGAGC 593  
 718 CAAAGATTGGGTATGCAATGGGTGGGCTGACAACTTGTCTTTTGGTGGTATCCAACT 777

RESULT 10  
 AAQ98578  
 ID AAQ98578 standard; DNA; 2045 BP.

QY 594 AAGGTGAC-ATCTTTGGCGA---GGCGGGAGCATGTCCGTTGTTGTCGCCACTCATCTGG 649  
 DB 778 AAGTTACTACTTTTGGTGAATCTGCTGGTCTATGTCCTGTCATGTGTCATTTTGTGG 837  
 QY 650 AACGACGGCAGCAACACGTATCAAGGGCAAGCCGTTGTTCCGCGGGGCGCATCATGCA---G 706  
 DB 838 AACGACGGTGAACAACCTTCAAGGGTAAGCCATTGTTTCAGAGCTGGTATCATGCAATCT 897  
 QY 707 GGAGCCATGTGC---CGGACCCGGTGGAGCGGACGTACGGCAACGAGATCTACGACCTC 763  
 DB 898 GGTGCTATGTTTCATCTGACCGCGTGCAGCGTATCTACGGTAACGAAATTTTTCAGCTTG 957  
 QY 764 TTTGTCTCGAGTGTGCTGTGGCAGCGCCAGCGACAAAGCTCGGTGCTTGGCGAGTG-- 821  
 DB 958 TTGGCTTCCAACGCTGGTGTGCTCTGCAAGTTGGCTTGTGAGAGGTGTT 1017  
 QY 822 -CGAGCGACCTTGTGCTGATGCCAACCAACACATCTCTGGGTTCTTGGGCTACTCTCG 880  
 DB 1018 TCTTCTGACACTTTGGAAGACGCCAACCAACACACCTCTGTTCTTGGCTTACTCTCTCC 1077  
 QY 881 TTGGGTTGTACT-----CCGCGCGCGAGCGGCAAGAACATCACCGATGATGATCAAG 934  
 DB 1078 TTAAGATTGTCTTACTTTGCCAAGACACGAGCGTGTAAACATCACCGACGATGTACGCT 1137  
 QY 935 TTGGTGGCGCAGCGCAAGTATGCAAGCTTCCGCTGATCATTTGGCGACCAAGACGAG 994  
 DB 1138 TTGGTTAGAGAGGTAAAGTATGCAATCCCTGTTATCATCGTGAACCAACGAGAA 1197  
 QY 995 GGACCATCTTTGGCTC-----TTGAACGTGACCAACGAAATGCTCAGGCGCGTGTAC 1047  
 DB 1198 GGTACCTTCTTGGTACTTCTTCTTTGAAGCTTACCACCTGATGCCAAGCCAGAGATAT 1257  
 QY 1048 TTCAAGCA---GTTCTATCAGCGCAGGAGCGGAGATCGACACTTGTGATGGCGGCTAC 1104  
 DB 1258 TTCAAGCAACTTCTTGTCCAGCTAGCGAGCTGAAATCGACACTTTGATGACTGCTTAC 1317  
 QY 1105 CCCAGGACATCACCGAGGT-----CGTTTCGACACGGTTCACGCTCACCCCG 1154  
 DB 1318 CAGGTGACATCACTCAAGTTTCCATTGACACTGGAATTTAAACGCTTTGACCCCA 1377  
 QY 1155 CAGTTCAAGAGATC---GCGGTGCTCGGCGACTTTGCAATTCATCCACGCCCGCGCTAC 1211  
 DB 1378 CAATTCAAGAGATCTCTGCTGTTTGGGTGACTTGGGTTTACTTTTGGTCTGTAGATAC 1437  
 QY 1212 TTCTCAACCACTTCCAGGGCGCACCAAGTACTCGTTCCTCAAGAGCTC-----GGG 1265  
 DB 1438 TTCTTGAACCACTACACCGGTGGTACCAAGTACTCTTCTTGTCTAAGCAATTTGCTGTT 1497  
 QY 1266 TTGCCAATCATGGGCACTTCCATGCCAAGCATTTGTGGCAGGACTACTTGTGTTGGGA 1325  
 DB 1498 TTGCCAGTTTGGGTACTTTCCACTCCAAGATATGCTTCTTCAAGACTACTTGTGTTGGGT 1557  
 QY 1326 AGCGGAGCGTCTATCTACAAACACGCGTTTATGCGGTTCGCCACCGACTTGGACCCCAAC 1385  
 DB 1558 TCTGGTTCTTGTATCTACAAACGCTTTCAATTTGCTTTTGGCAGTACTGTTGGACCAAC 1617  
 QY 1386 ACCGCGGGTGTGTGTTGAGACTGGCCCAAGTACAC-----AGCAGCCAGGGCAACAAAC 1439  
 DB 1618 ACCGCGGGTGTGTGTTGAGTGAAGTGGCCAGAAATACACCTCTTCTCTCAATCTGTTAAACAAAC 1677  
 QY 1440 TTGATGATCATCAACGCTTGGCTGTATACCGGCAAGCAACTTCCGACCGCTGGC 1499  
 DB 1678 TTGATGATCATCAACGCTTGGGTGTGATACCGGTAAAGCAACTTCAAGACCGCGCGGT 1737  
 QY 1500 TACGACCGCTTGTGATGACCAACCC 1522  
 DB 1738 TACGACCGCTTGTCTTCCAAACCC 1760

XX AAQ98578;  
 XX 16-OCT-2003 (revised)  
 DT 05-MAR-1996 (first entry)  
 XX  
 XX Aspergillus foetidus glucoamylase gene promoter.  
 DE  
 XX recombinant; glucose oxidase; GOD; extracellular production;  
 KW filamentous fungus; glucoamylase; promoter; ss.  
 KW  
 XX Aspergillus foetidus; SE4.  
 OS  
 XX EP665291-Al.  
 FN  
 XX 02-AUG-1995.  
 PD  
 XX 19-JAN-1995; 95EP-00200129.  
 PF  
 XX 28-JAN-1994; 94BE-00000102.  
 PR  
 PR 17-JUN-1994; 94BE-00000586.  
 PR 09-JAN-1995; 95BE-00000014.  
 XX  
 XX (SOLV ) SOLVAY & CIE.  
 PA  
 XX Carrez D, Roos J;  
 PI  
 XX WPI; 1995-264864/35.  
 DR  
 XX  
 XX Expression system for extracellular prodn. of glucose oxidase - related  
 PT vectors and transformed cells, esp. new strain of Aspergillus foetidus,  
 PT provides high yields of enzyme uncontaminated by catalase.  
 PT  
 XX Claim 18; Page 17-19; 32pp; French.  
 PS  
 XX The promoter and terminator sequences from the glucoamylase gene of  
 CC Aspergillus foetidus SE4 strain (AAQ98578 and AAQ98570, respectively) are  
 CC used for extracellular expression of glucose oxidase in A.foetidus hosts.  
 CC The specifically claimed transformant strain SE4tr contains a vector in  
 CC which the GOD gene is under transcriptional control of the new promoter  
 CC and terminator sequences; extracellular production of GOD by the  
 CC transformant was 5000-10000 times greater than for the untransformed  
 CC parental SE4 strain. (Updated on 16-OCT-2003 to standardise OS field)  
 CC  
 XX Sequence 2045 BP; 454 A; 538 C; 565 G; 488 T; 0 U; 0 Other;  
 SQ  
 Query Match 11.3%; Score 173; DB 2; Length 2045;  
 Best Local Similarity 56.6%; Pred. No. 5.2e-28;  
 Matches 426; Conservative 0; Mismatches 310; Indels 17; Gaps 5;  
 274 GAGGACTGCTCACCATCAACGTTGGTGGCGCGCGGACCAAGCGGGCGCCCAACCTC 333  
 DB 70 GAGGATTGCTGAACATTGACATTGCGGTGCGCGCGGACCAAGCGGACTCGAAGCTG 129  
 QY 334 CCGGTTCATGCTCTGATCTTTGGCGGTGGTTTGAGATCGGAGCCGCCACCATCTTCCCT 393  
 DB 130 CTGTGCTGCTGGATCTTTGGCGGAGCTTTGAACTTGGTTCAAAGCGGATGATGAT 189  
 QY 394 CCGGCCGAGATGGTCAACGAAGTGTGCTCATGGGCAAGCC-ATCATCAGCTGGCCGTC 452  
 DB 190 GGTAAACATGGTATCATCGTCGATAGACAAGAACATCGCTATCGTTTGTAGCAATG 249  
 QY 453 AACTACCGTGTGCTCGTGGGGTCTTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 512  
 DB 250 AATTATCGGTGGAGGTTTGGGTTTCTTGGCGGAAAGAGATCTTGAGAGCGGCTCC 309  
 QY 513 GGAACGCGGGTCTGAAGGACCAAGCGTTTGGGCATCCAGTGGTGGCAGCAACATTGCC 572  
 DB 310 GCGAAGCTAGGCTCTCTGACCAACGCCCTTGCCCTGAGTGGTGGTGGTGGTGGTGGTGG 369  
 QY 573 GGGTTGCGGGGACCCCGGAGAGGTGACATCTTTGGCGAGGGCGGCGAGCATCTCCGTGT 632  
 DB 370 GCCTTTGGTGGAGACCGGACCAAGGTGACGATTGGGGAGAAATCAGCAGGA-GCCATCG 428

QY 633 TGTGCCACCTCATCTGGAAACGAGCGGACAAACAGTAAAGGGCAAGCCGTTGTTCGCG 692  
 DB 429 TTTGACTAGATGACTTGTACGACGGAACATCACTTAAAGGATAAGCCCTTGTTCGCGG 488  
 QY 693 CGGCAATCATGACGAGGAGCCATGGTGC-----CGGACCGGTGGACGGCAGTACGGCA 746  
 DB 489 GGGCCATCATGGACTCCGGTAGTGTGTTCCCGCAGACCCCGTCGATGGGGTCAAGGAC 548  
 QY 747 ACAGATCTAGACCTCTTTGTCTCGAGTGTGCTGTGGCAGCGCCAGCGACCAAGTTCG 806  
 DB 549 AGCAAGTATATGATGCGGTAGTGAATCTGCAGGCTGTCTCTTCTTAACGACACCCCTAG 608  
 QY 807 CGTCTTGGCGAGTGGCAGGACAC---CTTGTCTGATGTCACCAACAACTCTCTGGGT 863  
 DB 609 CTCTGCTGGTGAACACTAGACTACACCGACTTCTCAATGCGGCAAACTCCGTGCCAGCA 668  
 QY 864 TCCTGGCGTACTCTCTGTTGGGTTGTACT-----CCCGGCCGACGCGGCAAGACATCA 917  
 DB 669 TTTTAAGCTACCATTTCTGTGGCGTTATCATATGTCCTCGACCGGACGGGCGGTGT 728  
 QY 918 CCGATGACATGTACAAGTTGGTGGCGGACGCAAGTATGCAAGCGTTCCCGTGCATATG 977  
 DB 729 CGGATCACCGACGTTTGGGCAAGCAGGGAATATGCTCGGCTCCCGTTCATCGTGG 788  
 QY 978 CGGACCAAGACGACGAGGCAACCATCTTTGGCT 1010  
 DB 789 CGGACCAAGAGATGAGGGGACCTTATTTCGCT 821  
 RESULT 11  
 AAT11299  
 ID AAT11299 standard; DNA; 2045 BP.  
 XX  
 AC AAT11299;  
 XX  
 DT 16-OCT-2003 (revised)  
 DT 25-JUL-1996 (first entry)  
 XX  
 DE Aspergillus foetidus SE4 glucoamylase promoter sequence.  
 XX Microbial aspartic protease; filamentous fungus; aspergillopepsin;  
 KW extracellular production; secretion; glucoamylase gene promoter; ss.  
 KW  
 XX Aspergillus foetidus; SE4.  
 OS  
 PN EP687734-Al.  
 XX  
 PD 20-DEC-1995.  
 XX  
 PF 09-JUN-1995; 95EP-00201521.  
 XX  
 PR 17-JUN-1994; 94BE-00000585.  
 PR 09-JAN-1995; 95BE-00000015.  
 XX  
 PA (SOLV ) SOLVAY SA.  
 XX  
 PI Carrez D, Dhaese P;  
 XX  
 DR WPI; 1996-031780/04.  
 XX  
 PT System for extracellular prodn. of microbial aspartic protease - contg.  
 PT promoter, signal sequence, pro-peptide and mature enzyme sequences, and  
 PT terminator, esp. for aspergillopepsin prodn.  
 XX  
 PS Claim 19; Page 18-20; 35pp; French.  
 XX  
 CC Spores of A.foetidus ATCC 14916 were exposed to UV, then grown on a  
 CC nutrient medium and one mutant having increased production of  
 CC glucoamylase was selected and designated SE4. The promoter and terminator  
 CC regions of the glucoamylase gene from mutant strain SE4 were isolated and  
 CC sequenced. These transcription control regions are preferred for  
 CC incorporation into novel expression vectors designed to allow

CC extracellular production of microbial aspartic protease. In particular, the enzyme aspergillopepsin is produced by transformed filamentous fungi and secreted into the culture medium due to the presence of a fungal secretion peptide. The present sequence is that of the A.foetidus S84 CC glucoamylase promoter. (Updated on 16-OCT-2003 to standardise OS field.)  
XX  
SQ Sequence 2045 BP; 454 A; 538 C; 565 G; 488 T; 0 U; 0 Other;

Query Match 11.3%; Score 173; DB 2; Length 2045;  
Best Local Similarity 56.6%; Pred. No. 5.2e-28;  
Matches 426; Conservative 0; Mismatches 310; Indels 17; Gaps 5;  
274 GAGGACTGCTCCACATCAAGTGGTGGCGCGCCGACCAAGGGCGGCGCAACCTC 333  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
70 GAGGATTGCTGAACATTGACATTCGGCGTCCGGCGGGACACCGCGGACTCGAAGCTG 129  
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
334 CCGGTATGCTCTGGATCTTTGGCGGTGGTTTGGATCGGAGCCCCACATCTTCCT 393  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
130 CCTGTGCTGCTGGATCTTTGGCGGAGGCTTTGAACCTTGGTTCAAGGGGATGTATGAT 189  
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
394 CCGGCCAGATGGTCAACAAAGTGTCTCATGGCAAGCC-ATCATCCACGTGGCGTC 452  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
190 GGTACACAGATGTATCATCGTATAGACAGACATGCTATCGTGTGTGATGATG 249  
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
453 AACTACCGTGTGCTGCTGGGGTCTTGGCTGGTATGACATCAAGGGCGGCGGAGC 512  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
250 AATTATCGCTGGAGGTTTCGGGTTCTTCCCGGAAAGAGATCCTGGAGGACGGGTCC 309  
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
513 GGAACCGCGCTTGAAGGACGAGCTTTGGCATGCGAGTGGTGGCAGACACATGCC 572  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
310 GCGAACTAGGGCTCTCTGACCAACGCTTTGCCCTGCGAGTGGTGGCGACACATCGAG 369  
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
573 GGGTTCCGGCGGACCGAGCAAGGTGACATCTTTGGCGAGGCGGCGCATGTCGTGT 632  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
370 GCCTTTGGTGGACCGGACAGGTGACATTTGGGGAATCAGCAGGA-GCCATTCG 428  
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
633 TGTGCACCTCATCTGGAACGACGCGGACAAACAGTCAAGGGCAAGCCGTTGTTCCGCG 692  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
429 TTTGACTAGATGACTTGTACGACGGAACATCATCTTACAAAGGATAAGCCCTTGTTCGGG 488  
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
693 CGGGCATCATGAGGAGCCATCGTGC-----CGGACCGGTGGACGCGCATCGGCA 746  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
489 GGGCCATCATGGAATCCCGTAGTGTGTTTCCCGCAGACCCCGTCGATGGGGTCAAGGGAC 548  
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
747 ACAGATCTACGACCTCTTTGCTCGAGTGTGCTGGCTGGCGAGCGGACGACAGCTCG 806  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
549 AGCAAGTATATGATGCGGTAGTGGATCTGAGGCTGTTCTTTTAAGCACACCCCTAG 608  
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
807 CGTGTTCGGCAGTGGAGCGACAC---CTTGTCTGATGCCAACCAACACACTCTCTGGGT 863  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
609 CTGTGCTGGGTGAACATAGACTACACCGACTTCTCAATGGGCAAACTCCGTGCCAGGCA 668  
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
864 TCTTGGCGTACTCTCTGTTGCGGTGTACT-----CCGGCCGAGGCGGAAGACATCA 917  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
669 TTTTAAGTACCAATCTGTGGCGTTTATCATATGCTCTCGACCGGACGGCGCGTGT 728  
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
918 CGGATGACATGTACAAGTTGGTGGCGGACGGAAGTATGCAAGCTTCCGTGATCATTTG 977  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
729 CGGCATCAGCGACGTTTGGGCAAGACAGGGAATATGTCGGTCCCGCTTCATCGTGG 788  
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
978 GCGACAGAACGACGAGGCGACCATCTTTGGCT 1010  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
789 GCGACCAAGAGGATGAGGGGACCTTATTCGCT 821

RESULT 12  
AAQ05605  
ID AAQ05605 standard; cDNA; 1674 BP.  
XX  
AC AAQ05605;  
XX  
DT 24-OCT-2003 (revised)  
DT 25-MAR-2003 (revised)

DT 13-DEC-1990 (first entry)  
XX Gene encoding protein with lipase activity.  
DE Lipase.  
XX Galactomyces geotrichum; ATCC34614.  
XX Key Location/Qualifiers  
FT misc\_RNA 40..1671  
FT /\*tag= a  
FT /label= Claim 1  
XX JP02174680-A.  
PN 06-JUL-1990.  
XX 27-DEC-1988; 88JP-00330598.  
XX 27-DEC-1988; 88JP-00330598.  
XX (KURK ) KURITA WATER IND LTD.  
PA (OSAK ) OSAKA CITY.  
XX WPI; 1990-250686/33.  
DR P-PSDB; AAR96370.  
XX Protein gene having lipase activity - has defined sequence of 544 bases and gives lipase producing vector on integration to expression (secretion) vector.  
XX Disclosure; Fig 4; 12pp; Japanese.  
XX The cloned cDNA can be inserted into an expression vector and used to transform hosts for the prodn. of a protein with lipase activity. See also AAQ05606. (Updated on 25-MAR-2003 to correct PA field.) (Updated on 24-OCT-2003 to standardise OS field)  
XX Sequence 1674 BP; 351 A; 473 C; 381 G; 469 T; 0 U; 0 Other;  
Query Match 11.2%; Score 171.2; DB 2; Length 1674;  
Best Local Similarity 52.8%; Pred. No. 1.2e-27;  
Matches 643; Conservative 0; Mismatches 508; Indels 67; Gaps 10;  
QY 270 GAGTAGGAGTCTCCTCACCATCAACCTGTGTGGCGCGCGGACCAAGCGGCGCCAA 329  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
342 GAATGAGGACTGTCTTTTACCTCAATGTTTTCGCGCTGTGGCACCAGCCTGATGCTAA 401  
QY 330 CTCTCCCGGTCTCTGGATCTTTGGCGGTGGTTTGAGATCGGCAGCCGCCCATCTT 389  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
402 GCTCCCGGTCTGTGTTTACGTTTACGTTGTGGTGTGTTTGTTCGTTCTCTGCTGCTA 461  
QY 390 CCTCTCCCGGTCTGTGTTTACGTTTACGTTGTGGTGTGTTTGTTCGTTCTCTGCTGCTA 448  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
462 CCTGTGTAACAGTACGTTTAAAGGAAAGTATCAACATGCGGCGGCGGCTGTGTTGTTTC 521  
QY 449 CGTCAACTACCGTGTGCTCTGGGGTCTTGGCTGTGTTGATGATCAAGCGCGGCGG 508  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
522 CATCAACTACCGTGTGCTCTGGGGTCTTGGATTTCTGGGTGTGATGATCATCCGCTGAGGG 581  
QY 509 CAGCGGGAACGCGGCTTGAAGGACGAGGCTTTGGGCGATGCGAGTGGTGGCAGACACAT 568  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
582 CAAACCAACGCTGTGCTGCACGACGACGAGGCTCTCGAGTGGTGGTTAGCGACACAT 641  
QY 569 TGCCTGGTTCGCGGCGACCCGAGCAAGGTGA----CATCTTTGGCGAGCGGCGAGCAT 624  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
642 TGCCAACTTTGGTGGTGTATCCCGATAAGGTATGATTTTCGGGTGAGTCCGCTGGTCCAT 701  
QY 625 GTCCGCTGTGTCACCTCATCTGGNACCGACGCGGCAACACATGACAAAGGCGGCTT 684  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
702 GAGTGTGCTCACCAGCTTATTCCTATGTTGGTGTGACAAACCTACACGGAAGAAGCT 761  
QY 685 GTTCCGCGCGGCGCATCATCAG---GGAGCCATGGTGGTGGCGGACCCCGGTGGACGCGCA 741

Db	762	TTTCCACTCTGCCATCTTCAGTCTGGTGGCCCTCTTCTTACCAAGACTCTAGTCGGT	821
Qy	742	CGGCAACGAGATCTACGAC-----CTCTTTGTCTCGAGTGTCTGTGGCAG	789
Db	822	TGGTCCCGATATTTCTTACACAGATTTGCTCAGTATGCCGGATGTGACACTAGTGGCAG	881
Qy	790	CGCCAGCGACAGTCTGGTCTGTCGCGAGTGGCGAGCGACACCTTGTCTGATGCCACAA	849
Db	882	TGCCAACGACACTCTGGAGTGTCTCCGCGAAGTCCAGCTCTGTCTTGACAGTCCCA	941
Qy	850	CAACACTCTCTGGGTTC-----TTGGGGTACTCC-----TGTGTGCGGTGTGATCT	894
Db	942	GAACTCTTACGATCTCAAGATCTGTTTGGTCTACTCCCTCAATTTCTTGGATTGGTCC	1001
Qy	895	CGGCGCCGACGGCAAGAACATCACCAGTACATGTAAGTGTGGCGCGAGCGCAAGTA	954
Db	1002	CAGACCCGACGCGCAACATTAATTCGCGATGCGCTTATGAGCTCTTCCGACGGTAGATA	1061
Qy	955	TGCAAGCGTTCCTGGTATCTTGGCGACAGAAACGAGCGGACCATCTTTGGCTCT--	1012
Db	1062	CGCAAGTTCCTTACATTAAGGTTAACGAGATGAGTACTGTCTTTGTCTCTGT	1121
Qy	1013	-----TGAACTGTACACGAATGCTCAGCGCGGTGCTTACTTCAAGCA-----GTTCAATCA	1064
Db	1122	TGCTCTCAACGCTACCAAGACTCCCATGTTAAGAGTGGTTGCAGTACATTTTCTACGA	1181
Qy	1065	CGCCAGCGCGGAGATGACACCTTGTATGGCGCGTACCCCGAGGACATCACCAGGG	1124
Db	1182	TGCTTCCGAGGCTTCCATTTGACCGTGTGTTGTGCTGTACCGCGACACCTCTCTGTGG	1241
Qy	1125	TCCG-----TTCCGACAGGTTCAACGCTCACCCGCGATTTCAAGAGAATCGCGT	1174
Db	1242	CTGCGCCTTCGCACTGGCAATCTTAATGCCCTGACCCCGCAAGTTCAAGCGTGTGGCG	1301
Qy	1175	GCTCGCGCACTTGCATTCATCCAGCGCGCGCTACTT-----CCTCAACA	1222
Db	1302	CATCTTGTCCGATATGCTTTTCCAGTCTCCCGCGCGTGTAGCTTAGCGCCACCAAGA	1361
Qy	1223	CTTCCAGGCGGCGACCAAGTACTGCTTCTCAGCAGTCTGGGTGGCAATCATGGCAC	1282
Db	1362	CGTTAACCGCTGGACTTACCTTTTCGACCCATCTGCACAACTGGTGCCATTTTGGGTAC	1421
Qy	1283	CTTCCATCCCAACGATTTGTGTGGCAGACTTGTGGAAGCGGAGCGTCACTTA	1342
Db	1422	TTTCCATGGCAAGAGTTCATCTTCCATTCATGTGACATTTGCGCGCTAACTCTCTA	1481
Qy	1343	CAACAACCGTTTATCGGTTTGGCCACCGACTTGAACCCCAACACCGGGGTTTGTGT	1402
Db	1482	CCTTCGTTATTTTATTTCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	1541
Qy	1403	GAACTGGCCCAAGTACAC	1420
Db	1542	CCAGTGGGATCAATACAC	1559
RESULT 13			
AAQ54020	AAQ54020 standard; DNA; 1828 BP.		
XX	AAQ54020;		
AC	AAQ54020;		
DT	16-OCT-2003 (revised)		
DT	25-MAR-2003 (revised)		
DT	19-JUL-1994 (first entry)		
XX	Lipase coding sequence of Geotrichum candidum.		
DE	Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;		
KW	Major cell wall protein; glycosyl-phosphatidyl-inositol;		
KW	anchoring protein; alpha factor; alpha-agglutinin; invertase; inulinase;		
KW	alpha-amylase; Geotrichum candidum; lipase; enzymatic process;		
KW	fermentation; biodegradation; catalysis; ss.		

XX	OS	Galactomyces geotrichum.
XX	Key	Location/Qualifiers
PH	CDS	40..1731
FT		/*tag= a
FT		/product= "Lipase."
FT	sig_peptide	40..96
FT		/*tag= b
FT	mat_peptide	97..1728
FT		/*tag= c
FT		/product= "Lipase."
XX	XX	WO9401567-A1.
PN	XX	20-JAN-1994.
PD	XX	07-JUL-1993; 93WO-EP001763.
XX	XX	08-JUL-1992; 92EP-00202080.
PR	14-DEC-1992; 92EP-00203899.	
XX	(UNIL ) UNILEVER PLC.	
PA	(UNIL ) UNILEVER NV.	
XX	Klis FM, Schreuder MP, Toschka H, Verrips CT;	
PI	WPI; 1994-035071/04.	
DR	P-PSDB; AAR47577.	
XX	Immobilisation of enzymes to microbial cell wall - by prodn. of fusion	
PT	protein of enzyme linked to anchoring protein.	
PS	Claim 8; Page 49-52; 99pp; English.	
XX	The lipase is used in a method to immobilise enzymes to a microbial cell	
CC	wall. The coding sequence is used in the production of a recombinant	
CC	polynucleotide which comprises a structural gene encoding a protein with	
CC	catalytic activity (the lipase) and at least part of a gene encoding at	
CC	least the C-terminus of a protein capable of anchoring in a eukaryotic or	
CC	prokaryotic cell wall. The anchoring fragment or protein is selected from	
CC	alpha agglutinin, AGA 1, FLO 1, major cell wall protein of lower	
CC	eukaryotes or a proteinase of lactic acid bacteria. The recombinant	
CC	polynucleotide preferably also comprises a sequence encoding a signal	
CC	peptide to ensure secretion of the expressed product. The signal peptide	
CC	is preferably derived from glycosyl-phosphatidyl-inositol, anchoring	
CC	protein, alpha factor, alpha-agglutinin, invertase or inulinase, alpha-	
CC	amylase of Bacillus or proteinases of lactic acid bacteria. The host	
CC	microorganism can be used for performing enzymatic processes on an	
CC	industrial scale. (Updated on 25-MAR-2003 to correct PN field.) (Updated	
CC	on 16-OCT-2003 to standardise OS field)	
XX	Sequence 1828 BP; 389 A; 483 C; 419 G; 537 T; 0 U; 0 Other;	
Qy	Query Match	10.6%; Score 162; DB 2; Length 1828;
Db	Best Local Similarity	56.0%; Pred. No. 1.3e-25;
XX	Matches 448; Conservative	0; Mismatches 315; Indels 37; Gaps 6;
Qy	248	TGTTCCAGCGGTGCTTCCCGAGTGGAGTCTCTCACCATCAACGTTGTCGCGCCGC 307
Db	377	TGGCGAGGTAGTGTCTCCATGAATGAGACTGTCTACCTTAACGTTTCCGCGCCG 436
Qy	308	CGGCGACCAAGCGCGCGCCACCTCCCGGTCTATGCTCTGAGATCTTGGCGGTGGTGG 367
Db	437	CTGGCACCAGCGCTGATGCTAGCTCCCGTCTGATGTTGAGTTTACGTTGCTGCTTG 496
Qy	368	AGATCGGAGCGCCACCATCTTCCCTCCCGCCAGATGTCACAGAGTGTCTCATGG 427
Db	497	TGTTTGGTCTTCTGCTTCTTACCTTGGTACCGGTACGCGGTACGAGAGTGTGGAAATGG 556
Qy	428	GCAAGCC-ATCATCCAGTGGCGCGTCAACTACCGTGTTCCTCTGTCGGGGTCTTGGCTG 486
Db	557	GCCAGCGTGTGTTGTTTTCATCACTACCGTACCGGCCCTATGATGATCTTGGGTG 616

Query Match	10.6%;	Score 162;	DB 2;	Length 1828;
Best Local Similarity	56.0%;	Pred. No. 1.3e-25;		
Mismatches	448;	Conservative	0;	Mismatches 315; Indels 37; Gaps 6;







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2004, 15:47:40 ; Search time 95 Seconds  
(without alignments)  
8949.313 Million cell updates/sec

Title: US-09-943-857-3  
Perfect score: 1532  
Sequence: 1 gccaccgcaagctgcga.....tgaccaaccgtctttgtg 1532

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 582709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfileal.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1022	66.7	1650	2	US-08-776-210-6
2	173	11.3	2045	1	US-08-379-926A-7
3	164.2	10.7	1828	3	US-08-362-525-11
4	161.8	10.6	1635	2	US-08-776-210-5
5	88.2	5.8	1738	2	US-08-932-376A-1
6	87	5.7	8533	1	US-07-846-181-6
7	87	5.7	8533	1	US-07-845-989-6
8	61.2	4.0	2184	1	US-08-445-050-8
9	61.2	4.0	2184	1	US-08-204-691-8
10	61.2	4.0	2428	1	US-08-445-050-1
11	61.2	4.0	2428	1	US-08-204-691-1
12	61.2	4.0	2428	4	US-09-355-295B-2
13	61.2	4.0	2487	3	US-08-370-223-12
14	60	3.9	1845	1	US-07-732-962A-1
15	60	3.9	1845	5	PCI-US92-06106-1
16	60	3.9	2256	2	US-08-318-826A-5
17	60	3.9	2256	2	US-08-370-156-1
18	60	3.9	2256	3	US-08-814-095-1
19	60	3.9	3016	2	US-08-318-826A-7
20	60	3.9	3016	2	US-08-370-156-5
21	60	3.9	3016	3	US-08-814-095-5
22	60	3.9	3096	2	US-08-318-826A-6
23	60	3.9	3096	2	US-08-370-156-3
24	60	3.9	3096	3	US-08-814-095-3
25	60	3.9	35060	3	US-08-814-095-7
26	59.6	3.9	2344	4	US-09-347-878-31
27	59.6	3.9	3018	1	US-08-347-718B-3

28 59.6 3.9 3018 1 US-08-482-262-3 Sequence 3, Appli  
29 59.6 3.9 3018 6 5200183-1 Patent No. 5200183  
30 59 3.9 2220 2 US-08-932-376A-3 Sequence 3, Appli  
31 56.6 3.7 1905 4 US-09-347-878-33 Sequence 3, Appli  
32 56.6 3.7 1907 1 US-08-462-884A-2 Sequence 2, Appli  
33 56.6 3.7 1908 1 US-08-461-881B-2 Sequence 2, Appli  
34 56.6 3.7 1908 2 US-09-123-960-2 Sequence 2, Appli  
35 51.2 3.3 1746 4 US-10-023-515-3 Sequence 3, Appli  
36 51.2 3.3 2158 4 US-10-023-515-1 Sequence 1, Appli  
37 49 3.2 4403765 3 US-09-103-840A-2 Sequence 2, Appli  
38 49 3.2 4411529 3 US-09-103-840A-1 Sequence 1, Appli  
39 47.8 3.1 1590 4 US-09-351-150A-2 Sequence 2, Appli  
40 47.8 3.1 1800 1 US-08-484-815-11 Sequence 11, Appli  
41 47.8 3.1 1800 3 US-08-888-949-11 Sequence 11, Appli  
42 47.8 3.1 1800 3 US-08-888-950-11 Sequence 11, Appli  
43 47.8 3.1 1800 3 US-09-262-758-11 Sequence 11, Appli  
44 47.8 3.1 1800 4 US-09-885-876-11 Sequence 11, Appli  
45 47.8 3.1 1800 4 US-09-885-901-11 Sequence 11, Appli

#### ALIGNMENTS

RESULT 1  
US-08-776-210-6  
; Sequence 6, Application US/08776210  
; Patent No. 5942659  
; GENERAL INFORMATION:  
; APPLICANT: ALIBERT, Gilbert  
; APPLICANT: MOULOUNGUI, Zephirin  
; APPLICANT: BOUDET, Alain  
; TITLE OF INVENTION: PROCESS FOR PRODUCING FATTY ACIDS OR  
; DERIVATIVES THEREOF FROM OLEAGINOUS PLANTS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: YOUNG & THOMPSON  
; STREET: 745 South 23rd Street  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/776,210  
; FILING DATE: 24-JAN-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 94 09272  
; FILING DATE: 25-JUL-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/FR95/00957  
; FILING DATE: 18-JUL-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PATCH, Andrew J  
; REGISTRATION NUMBER: 32,925  
; REFERENCE/DOCKET NUMBER: IN 387 - BE 6996  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-521-2297  
; TELEFAX: 703-685-0573  
; TELEX: 248425 EMBON  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1650 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-776-210-6

Query Match	66.7%;	Score 1022;	DB 2;	Length 1650;
Best Local Similarity	84.7%;	Pred. No. 2e-241;		
Matches 1342;	Conservative	Mismatches 180;	Indels 62;	Gaps 15;
QY	2	CCACGCCCAAGCTCGCCAAACGGCGACACATCACGGGTCTCAACGCCCATCATCAACGAG	61	
DB	49	CCACGCCCAAGCTCGCCAAACGGCGACACATCACGGGTCTCAACGCCCATCATCAACGAG	108	
QY	62	GCCTTCCTCGCATTCCTTTTGCAGAGCGCGGTGGCAACCTCCGTTCAAGGACCCCT	121	
DB	109	GCCTTCCTCGCATTCCTTTTGCAGAGCGCGGTGGCAACCTCCGTTCAAGGACCCCT	168	
QY	122	GTCCGCTACTCTGGCTCGCTCAACGGCCAGAAGTT-----ACTTACGGCCCGGTGATG	174	
DB	169	GTCCGCTACTCTGGCTCGCTCATGTGACCAAGAGTTTCAACGCTGTACGGCCCGGTGATG	228	
QY	175	CACGAGAACCCGAGGGGACGTTTGAAGAGAACCTTGGCAAGACGGCACTCGACTTTGGTG	234	
DB	229	CACGAGAACCCGAGGGGACCTTACGAGGAGAACCTCCCAAGGAGAGCGCTCGACTTTGGTG	288	
QY	235	ATCAGTTCGAAGGTGTTTCCAGGCGGTCTTCCCGAGAGTGAGGACTGCTCTCAATCAAC	294	
DB	289	ATCAGTTCGAAGGTGTTTCCAGGCGGTCTTCCCGAGAGTGAGGACTGCTCTCAATCAAC	348	
QY	295	GTGGTGGCGCGCGGGGACCAAGAGGGGGGCGCAACTCCCGGTCTGCTGATCTTTT	354	
DB	349	GTGGTGGCGCGCGGGGACCAAGAGGGGGTGCCAACTCCCGGTCTGCTGATCTTTT	408	
QY	355	GGCGTGGGTTCAGATCGGACGCCACCATCTTCCCTCCCGCCAGATGCTCAACCAAG	414	
DB	409	GGCGGGCGGTTCAGATCGGAGGGGGGCGCAACTTCCCTCCCGCCAGATGCTCAACCAAG	468	
QY	415	AGTGTGCTCATGGCAAG-CATCATCCAGTGGCCGTCAACTACGCTGTTCCTCGTGG	473	
DB	469	AGCATTCGCATGGCAAGGCCATCATCCAGTGGCGTCAACTACGCGGTGCTCGTGG	528	
QY	474	GGGTTCCTGGCTGGTGATGATCATCAAGCCGAGGCGAGGGGAGACCGCGGCTTGAAGAC	533	
DB	529	GGGTTCCTGGCTGGCGACGAGATCAAGGGCGAGGGGAGTGCCTAACCGGTTTGAAGGAC	588	
QY	534	CAGGTTTGGCATGAGTGGGTGGGAGACAACTTGGCGGTTTCGGCGGCGACCCGAGC	593	
DB	589	CAGGCTTGGCATGAGTGGGTGGGAGACAACTTGGCGGTTTCGGCGGCGACCCGAGC	648	
QY	594	AAGGTGA-CATCTTTTGGCGA---GGCGGGGAGCATGTTCGTTGTGCCACCTCATCTGG	649	
DB	649	AAGGTGACATCTTTTGGCGAGCTGGCGGGGAGCATGTTCGTTGTGCCACCTCATCTGG	708	
QY	650	AACGACGGCGCAACAACGTACAGGGGAGCGGTTTTCGGCGGGGAGCATATGCA---G	706	
DB	709	AACGACGGCGCAACAACGTACAGGGGAGCGGCTCTTTCGGCGGGGAGCATATGAGCTG	768	
QY	707	GGAGCCATGGTGC---GGACCCGGTGGAGGGGACGTTACGGCAACGAGATCTACGACCTC	763	
DB	769	GGGGCCATGGTGGCGGCTGGACCGCGTGGAGCGCGGAGCGGAGCAAGCTTTCGCGGTG	828	
QY	764	TTTGTCTCGAGTGTGGCTGTGGCGAGCGGCGAGGACAAAGCTTCGCTGTTCGCGAGTGC-	822	
DB	829	TTGGGTGCAACGGGGGCTGGCGAGCGGCGGAGCGGAGCAAGCTTTCGCGGTG	888	
QY	823	--GAGCGACACTTGTCTGATGCGCAACAACTCTCTGGTTCCTGGGCTACTCTCTG	880	
DB	889	CTGAGCGACAAGTTTGGAGGAGCGCACCAACCAACCCCTGGGTTCTTTGGCGTACTCTCTG	948	
QY	881	TTGGGTTG-----TACTCCCGCCGAGCGGAGAGAACATCAACGATGACATGTACAAG	934	
DB	949	TTGGGTTGCTGTATCTTCCCGGGCGGAGCGGCTGGAATCAACGAGACATGTACGCC	1008	
QY	935	TTGGTGGCGGACGGCAAGTATGCAAGCGTTCCTCGTGTATTTGGCGGACCGAGACGAG	994	
DB	1009	TTGGTGGCGGAGGAGATGACCAACATCTCTGTATCATCGCGCGGAGACGAGGAG	1068	
QY	995	GGCACCATCTTTG-----GCTCTTGAAGGTGACCGAGAAATGCTAGGGCGGTGCTTAC	1047	

;; FILING DATE: 17-JUN-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: BE 09500014  
;; FILING DATE: 09-JAN-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: OBLON, NORMAN F  
;; REGISTRATION NUMBER: 24,618  
;; REFERENCE/DOCKET NUMBER: 3987-13-0  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 703-413-3000  
;; TELEFAX: 703-413-2220  
;; TELEX: 248855 OPAT UR  
;; INFORMATION FOR SEQ ID NO: 7:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2045 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; US-08-379-926A-7

Query Match 11.3%; Score 173; DB 1; Length 2045;  
Best Local Similarity 56.6%; Pred. No. 4 Se-33;  
Matches 426; Conservative 0; Mismatches 310; Indels 17; Gaps 5;

QY 274 GAGGACTGCTCACCATCAACATGCTGGTGGCGCCGCGGACCAAGGCGGCGCAACCTC 333  
|||  
DB 70 GAGGATTGCTGAACATTGACATTCGGCGTCCGCGCGGACCAACCGGACTCGAAGCTG 129  
|||

QY 334 CCGGTGATGCTGCTGATCTTTGGGCTGGTGGTGGATCGGAGCCGCCACATCTCCCT 393  
|||  
DB 130 CCGTGTGCTGCTGATCTTTGGGAGGCTTTGAACCTTGTTCAAGGCGATGATGAT 189  
|||

QY 394 CCGCGCCAGATGCTCAACAGAGTGTGCTCATGGGCAAGCC-ATCATCCACGTGGCGTC 452  
|||  
DB 190 GGTACAACGATGATCATCTGCTGATAGACAAGACATGCTATGCTTTGATGAATG 249  
|||

QY 453 AACTACGCTGCTGCTGCTGGGGTCTTTGGTGGTGGATGACATCAAGGCGGAGGCGAGC 512  
|||  
DB 250 AATTATCGCTGGAGGCTTTGGGCTTTTCCCGGAAAGAGATCCTGGAGGACGGTCC 309  
|||

QY 513 GGGAAACCGGCTTGAAGGACCAAGCTTTGGGATGAGTGGTGGTGGGATGAGGAGGCGAGC 572  
|||  
DB 310 GGAACCTTAGGCTCTCGGACCAACGCTTGGCCCTGAGTGGTGGTGGGATGAGGAGG 369  
|||

QY 573 GGGTTCGGGCGGACCGGAGCAAGTGTGATCTTTGGCGAGGCGGCGAGCATGCTCGGTG 632  
|||  
DB 370 GCTTTTGGTGGAGCCCGGACAGGTGACGATTTGGGGAGATCAGCAGGA-CCCATTCG 428  
|||

QY 633 TGTGCCACCTCATCTGGAACGAGCGGCAACACGTAAGGCGCAAGCCGTTGTTCCGGG 692  
|||  
DB 429 TTGACTAGATGACTTGTACGACGGAACATCACTTACAAGGATAAGCCCTTGTTCGGG 488  
|||

QY 693 CGGGCATCATGAGGAGCCATGCTG- - - - -CGGACCGGTGGAGCGCAGTACGGCA 746  
|||  
DB 489 GGGCATCATGAGTCCGCTAGTGTGTTTCCCGCAGACCCCGTGGTGGTTCAGGGAG 548  
|||

QY 747 ACGAGATCTACGACCTCTTTGCTCGAGTGTGGCTGTGGCGGCGGCGAGCAAGCTCG 806  
|||  
DB 549 AGCAAGTATATGTCGGTAGTGGATCTGAGGCTGTTCTTCTTAACGACACCCCTAG 608  
|||

QY 807 CGTGTTCGGAGTGGAGCGGAC- - - - -CTTGCTGATGCCACCAACACTCTCTGGGT 863  
|||  
DB 609 CTGTCTGGTGAACCTAGACTACACCGACTTCTCAATGGGCAAACTCCGTGCCAGGCA 668  
|||

QY 864 TCTTGGGCTACCTCTGTTGGGCTTGTACT- - - - -CCGGCCCGGCGGCAAGACATCA 917  
|||  
DB 669 TTTTAAGCTACCATCTGTGGCTGATCATATATGTCCTCGACCGGAGCGGCGGTGT 728  
|||

QY 918 CCGATGACATGTACAGTTGGTGGCGGCGGCAAGTATCAAGCGTTCCTCGTATCATG 977  
|||  
DB 729 CGGCATCCGAGCGTTTGGGCAAGAGGGAATATGCTGGGTCCCGTTCATCGTGG 788  
|||

QY 978 GCGACCAAGACGACGAGGCGACCATCTTTGGCT 1010  
|||  
DB 789 GCGACCAAGAGATGAGGCGACCTTATTCGCT 821  
|||

RESULT 3  
US-08-362-525-11  
; Sequence 11, Application US/08362525  
; Patent No. 6027910  
; GENERAL INFORMATION:  
; APPLICANT: KLIS, FRANCISCUS M.  
; APPLICANT: SCHREUDER, MAARTEN P.  
; APPLICANT: TOSCHKA, HOLSER Y.  
; APPLICANT: VERRIPS, CORNELIS T.  
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE  
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION  
; NUMBER OF INVENTION: PROTEIN  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARRY & CUSHMAN, L.L.P.  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington  
; STATE: D. C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,525  
; FILING DATE: 04-JAN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 92202080.5  
; FILING DATE: 08-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 92203899.7  
; FILING DATE: 14-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP93/01763  
; FILING DATE: 07-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KOKULIS, PAUL N.  
; REGISTRATION NUMBER: 16,773  
; REFERENCE/DOCKET NUMBER: 213289/T7020(V)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1828 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA.  
; ORIGINAL SOURCE:  
; ORGANISM: Geotrichum candidum  
; STRAIN: CMICC 335426  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 40..1731  
; OTHER INFORMATION: /product= "lipase"  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 40..96  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 97..1728  
; OTHER INFORMATION: /product= "lipase"  
; OTHER INFORMATION: /gene= "lipB"

US-08-362-525-11

Query Match 10.7%; Score 164.2; DB 3; Length 1828;  
Best Local Similarity 54.6%; Pred. No. 6.3e-31;  
Matches 436; Conservative 0; Mismatches 328; Indels 35; Gaps 4;  
  
QY 248 TGTTCAGGCGGTCTTCCACAGTGCAGACTGCTCCACCATCAAGTGTGGCGCGC 307  
DB 377 TGGCCAGGGTAGTGTCTCCATGAATGAGGACTGTCTACCTTAAGTGTTCGCGCGC 436  
  
QY 308 CGGCACCAAGCGGGCGCAACCTCCCGGTGATGCTCTGATCTTTGGCGGTGGTTG 367  
DB 437 CTGGACCAAGCTGATGCTAAGCTCCCGGTGATGCTTGGATTTACGGTGTGCTT 496  
  
QY 368 AGATCGGAGCCCAACATCTTCCCTCCCGCCAGATGCTTCAACAGTGTGCTATGG 427  
DB 497 TGTGTGTTCTTCTGCTTCTTACCTCGTACCGCTAGCTCAAGGAGAGTGTGAAATGG 556  
  
QY 428 GCAGCC-ATCATCCAGTGGCGCTCAACTACCGTGTGCTCGTGGGGTCTTGGCTG 486  
DB 557 GCCAGCTGTGTGTTGTTTCCATCACTACCGTACCGGCCCTATGATTTCTTGGTG 616  
  
QY 487 GTGATGATCATCAAGCGCGGAGCGGAAACCGCGCTTGAAGAACAGCGTTTGGGCA 546  
DB 617 GTGATGCCATCACCGCTGAGGGAACACCAACGCTGTCTGACGACGAGCGCAAGGTC 676  
  
QY 547 TGCAGTGGTGGCAGACAACTTGGCGGTTTGGCGGCGACCCGAGCAAGTGA-CA 602  
DB 677 TCGAGTGGGTTAGCGACAACTTGGTGGTGTGCTCCGACAAAGTGTATGATTT 736  
  
QY 603 TCTTGGCGAGCGGCGAGCATGCTCGGTGTTGTGCGACCTCATCTGGAACGACGCGACA 662  
DB 737 TCGGTGAGTCCGCTGTGTCCTCAAGTGTGTCTACCACTGTTGCTTACGGTGTGACA 796  
  
QY 663 ACACGTACAGGCAAGCGGTTGTTCCGCGGCGCATCATGCGAGGAG----- 710  
DB 797 ACACGTACAGGAAAGCAGCTTTTCCACTCTGCGATTTCTTCACTGCGCGCTCTCTTC 856  
  
QY 711 ---CCATGGTGGGACCCGGTGGAGCGGACGTACGGCAACGAGATCTACGACCTTTTG 767  
DB 857 CTACTTTGACTCTACTTCTGTGTGTCGCGAGTGCCTACAGAGATTGCTCAGTATG 916  
  
QY 768 TCTGAGTGTGGCTGTGGCAGCGCGCAGCAAGCTGCGGTGCTTGGCGAGTGGGAGG 827  
DB 917 CGGATGAGACCAAGTCCAGTGAATGATGACTCTGCTGTGTCTCCGACCAAGTCCA 976  
  
QY 828 ACACCTTGTGATGCCAACCAACATCTCTGGGTTCTTGGGCTACTCTCTGTTGC--- 884  
DB 977 GCGATGCTTGACAGTGGCGAAGTCTGATGATCTTAAAGGACCTGTTTGTCTGCTCC 1036  
  
QY 885 -----GGTTGTACTCCGCGCGCGAGCAAGCAATCAACGATGATGATCA 932  
DB 1037 CTCAATTCCTTGGATTTGGTCCGACCGCGAGCAACATTTTCCGATGCCGTTATG 1096  
  
QY 933 AGTTGGTGGCGACGCGCAAGTATGCAAGGTTTCCCGTGTATCATTTGCGCACCAAGAGCG 992  
DB 1097 AGCTTACCGAGCGGTAGATAGCAAGTTCCCTTACATTACTTGGCAACGAGGAGTG 1156  
  
QY 993 AGGCAACCATCTTTGGCTC 1011  
DB 1157 AGGGTACTATTCTTGCCCC 1175

RESULT 4  
US-08-776-210-5  
; Sequence 5, Application US/08776210  
; Patent No. 5942659  
; GENERAL INFORMATION:  
; APPLICANT: ALIBERT, Gilbert  
; APPLICANT: MOULOUNGUI, Zephirin  
; APPLICANT: BOUDET, Alain  
; TITLE OF INVENTION: PROCESS FOR PRODUCING FATTY ACIDS OR  
; TITLE OF INVENTION: DERIVATIVES THEREOF FROM OLEAGINOUS PLANTS

NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YOUNG & THOMPSON  
STREET: 745 South 23rd Street  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/776,210  
FILING DATE: 24-JAN-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 94 09272  
FILING DATE: 25-JUL-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR95/00957  
FILING DATE: 18-JUL-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: IN 387 - BE 6996  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-521-2297  
TELEFAX: 703-685-0573  
TELEX: 248425 EMBON  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1635 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-08-776-210-5

Query Match 10.6%; Score 161.8; DB 2; Length 1635;  
Best Local Similarity 55.9%; Pred. No. 2.3e-30;  
Matches 447; Conservative 0; Mismatches 317; Indels 35; Gaps 6;  
  
QY 248 TGTTCAGGCGGTGCTTCCACAGTGCAGACTGCTCCACCATCAAGTGTGGCGCGC 307  
DB 281 TGGCCAGGGTAGTGTCTCCATGAATGAGGACTGTCTACCTTAACGTTTTCCGCGCTG 340  
  
QY 308 CGGCACCAAGCGGGCGGCAACCTCCCGGTGATGCTCTGATCTTTGGCGGTGGTTG 367  
DB 341 CTGGACCAAGCTGATGCTAAGCTCCCGCTCATGTTGGATTTACGTTGGTGGCTTG 400  
  
QY 368 AGATCGGACGCGCCACCATCTTCCCTCCCGCCAGATGCTCAACAGAGTGTGCTCATGG 427  
DB 401 TGTGTTGTTCTTCTGCTTCTTACCTTGTAAACGGCTACGTAAGGAGAGTGTGAAATGG 460  
  
QY 428 GCAAGCC-ATCATCAAGTGGCGGCTCAACTACCGTGTGCTTCCCTGCGGGGTTCTTGGCTG 486  
DB 461 GCCAGCTGTGTTGTTTCTTCCATCACTACCGTACCGGCCCTATGGAATTCCTGGGTG 520  
  
QY 487 GTGATGATCATCAAGCGCGGAGCGGAAACCGCGGCTTGAAGGACCAAGCGTTTGGGCA 546  
DB 521 GTGATGCCATCACCGCTGAGGGTACACCAACGCTGCTGTGACGACCAAGCGCAAGGTC 580  
  
QY 547 TGCAGTGGTGGCAGACAACTTGGCGGGTTCGGGGCGGACCCGAGCAAGTGA-CA 602  
DB 581 TCGAGTGGGTAGCGACAACTTTGGTGTGATCCCGACAAAGTGTATGATTT 640  
  
QY 603 TCTTGGCGAGGGCGGCGACATGTCGTTGTGCGACCTCATCTCTGGAAGCGGCGACA 662  
DB 641 TCGGTGAGTCCGCTGCTGTCATGAGTGTGTACCAAGCTGTGTGCTTACGGTGGTGA 700  
  
QY 663 ACACGTACAGGCGCAAGCGGTTGTTCCGCGCGGCGCATCATGCAAGGAG----- 710





```

QY 808 GTGCTCGCAGTCGCGAGCGACAC---CTTGCTCGATGCCAACCAACACATCTCTGGGTT 864
Db 547 TTGTCTCGTGAGCTAGACGACTACACGACTATCTCAATGCGGCAAACTCGTCCGGGATC 606
QY 865 CTTGGCGTACTCTCTGTTGGGTTGTACTCCCGGCCGACGGCAAGAACATCACCAGTGA 924
Db 607 CTAGTTATCACCGTGGGCTATCATATGTGCTCGACACGGGACGGGCAATTTGTGCG 666
QY 925 CAT--GTACAAAGTTGGTGGCGGACGCAAGTATGCAAGCGTTCCCGTGATCATTTGGCGAC 982
Db 667 CGTCGCCAGATTTTGGGTAAAGCAGGGAAGTATGCGGGGTCCCATTCATCTGTTGGGCGAC 726
QY 983 CAGAACGACGAGGACCATCTTTGGCTTTGGAACGTGACACAGA 1027
Db 727 CAAGAGGATAGGGGACCTTGTTCGCTTGTTCAGTCTTACGA 771

```

## RESULT 8

```

US-08-445-050-8
; Sequence 8, Application US/08445050
; Patent No. 5763739
; GENERAL INFORMATION:
; APPLICANT: Blaeckberg, Lars
; APPLICANT: Edlund, Michael
; APPLICANT: Hansson, Lennart
; APPLICANT: Hernell, Olle
; APPLICANT: Lundberg, Lennart
; APPLICANT: Stromqvist, Mats
; APPLICANT: Toernell, Jan
; TITLE OF INVENTION: No. 5763739el Polypeptides
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,050
; FILING DATE:
; CLASSIFICATION: 800
; PRIORITY INFORMATION:
; APPLICATION NUMBER: US/08/204,691
; FILING DATE:
; APPLICATION NUMBER: SE 9300686-4
; FILING DATE: 01-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9300722-7
; FILING DATE: 04-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Steiner Ph.D., Richard J
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-850
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)819-8783
; TELEFAX: (212)354-8113
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2184 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens

```

```

; TISSUE TYPE: mammary gland
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 82..2088
; OTHER INFORMATION: /label= Variant_T
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 151..2085
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 1756..2052
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 1756..1788
; FEATURE:
; NAME/KEY: repeat_unit
; LOCATION: 1789..1821
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 1822..1854
; FEATURE:
; NAME/KEY: repeat_unit
; LOCATION: 1855..1887
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 1888..1920
; FEATURE:
; NAME/KEY: repeat_unit
; LOCATION: 1921..1953
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 1954..1986
; FEATURE:
; NAME/KEY: repeat_unit
; LOCATION: 1987..2019
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 2020..2052
; US-08-445-050-8

```

Query Match 4.0%; Score 61.2; DB 1; Length 2184;

Best Local Similarity 52.7%; Pred. No. 1.3e-05;

Matches 214; Conservative 0; Mismatches 173; Indels 19; Gaps 3;

```

QY 236 TGCAGTCCAAAGTGTTCAGGCGGTGCTTCCCGAGAGTGGAGTGGCTCCCATCAACATCAACG 295
Db 344 TGCAGGCCACCATCACCCAGGACAGCACCTACGGGATGAAGACTGCTGTACTCAACA 403

```

```

QY 296 TGGTGGCGCGCGCGGACCAAGCGGGCGCC---AACCTCCCGGTCAATGCTCTGGATCT 352
Db 404 TTGGGTGCCCCAGGGCAGAGCAAGTCTCCGGGACCTGCGCGTATGATCTGGATCT 463

```

```

QY 353 TTGGCGGTGGGTTTGAGATCGGACG-----CCACCATCTTCCTCCCGCCAG 402
Db 464 ATGGAGCGCCTTCTCTCATGGGTCCGGCCATGGGGCCAACTTCTCAACAACATCTCTGT 523

```

```

QY 403 ATGGTCACCAAGAGTGTCTCATGGCAAGCCATCATCCACGTGGCCGTCAACTACCGTG 462
Db 524 ATGACGGCGAGGAGATCGCCACACCGGAAACGTATCGTGGTCACTTCAACTACCGTG 583

```

```

QY 463 TTGCCTCGTGGGGTTCTTGGGTGTGATGACATCAAGCGCGAGGCGACGCGGAACGCG 522
Db 584 TCGGCCCCCTTGGGTTCTCAGCATCTGGGAC-----GCCAATCTGCCAGGTATATG 637

```

```

QY 523 GCTTGAAGAACAGCGTTTGGGCATGCAAGTGGGTGGCAGACAAACATTCGCGGTTTGGCG 582
Db 638 GCCTTCGGGATCAGCACATGGCCATTGCTTGGTGAAGAGGAATATCGCGGCTTTCGGGG 697

```

```

QY 583 GCGACCGCAGCAAGGTGACATCTTTGGCGAGCGCGGACGATGTCC 628
Db 698 GGGACCCCAACAACATCACGCTCTTCGGGGAGTCTGCTGGAGTGC 743

```

```

RESULT 9
US-08-204-691-8
; Sequence 8, Application US/08204691
; Patent No. 5827683
; GENERAL INFORMATION:
; APPLICANT: Blaeckberg, Lars
; APPLICANT: Edlund, Michael
; APPLICANT: Hansson, Lennart
; APPLICANT: Herneli, Olle
; APPLICANT: Lundberg, Lennart
; APPLICANT: Stroomqvist, Mats
; APPLICANT: Toernell, Jan
; TITLE OF INVENTION: No. 5827683el Polypeptides
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/204,691
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9300686-4
; FILING DATE: 01-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9300722-7
; FILING DATE: 04-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner Ph.D., Richard J
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-850
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)819-8783
; TELEFAX: (212)354-8113
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2184 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: mammary gland
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 82..2088
; OTHER INFORMATION: /label= Variant_T
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 151..2085
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 1756..2052
; FEATURE:
; NAME/KEY: repeat_unit
; LOCATION: 1756..1788
; FEATURE:
; NAME/KEY: repeat_unit
; LOCATION: 1789..1821
; FEATURE:
; NAME/KEY: repeat_unit

```

/	LOCATION:	1822...1854
/	FEATURE:	
/	NAME/KEY:	repeat_unit
/	LOCATION:	1855..1887
/	FEATURE:	
/	NAME/KEY:	repeat_unit
/	LOCATION:	1888..1920
/	FEATURE:	
/	NAME/KEY:	repeat_unit
/	LOCATION:	1921..1953
/	FEATURE:	
/	NAME/KEY:	repeat_unit
/	LOCATION:	1954..1986
/	FEATURE:	
/	NAME/KEY:	repeat_unit
/	LOCATION:	1987..2019
/	FEATURE:	
/	NAME/KEY:	repeat_unit
/	LOCATION:	2020..2052
/	US-08-204-691-8	
Query Match            4.0%; Score 61.2; DB 1; Length 2184;		
Best Local Similarity 52.7%; Pred. No. 1.3e-05;		
Matches 214; Conservative 0; Mismatches 173; Indels 19; Gaps 3;		
Qy	236	TGCAGTCCAAAGTGTTCCAGGCGGTCTTCCCCAGAGTGAGGACTGCTCACCATCAACG 295
Dd	344	TGAGGCCCATCATCCCAGGACGACCTAACGGGATGAAGACTGCCTGTACCTCAACA 403
Qy	296	TGTTGGCGGCGCGGGGACCAAGCGGGCGCC---AACCTCCGGTCATGCTCTGGATCT 352
Dd	404	TTTTGGGTGCCCGGAGGAGCAAGTCTCCCGGGACCTGCCCGTTTATGATCTGGATCT 463
Qy	353	TTGCGGTGGGTTTGAGATCGGCAGC-----CCCAACATCTTCCCTCCGCGCCAG 402
Dd	464	ATGAGGCGGCTTCTCTATGGGTCCGGCCATGGGGCCAATTCTCTCAACAACCTACCTGT 523
Qy	403	ATGTCAACCAAGAGTGTCTCATGGGCAAGCCATCATCCAGTGGCGGTCAACTACCGTG 462
Dd	524	ATGACGCGAGGAGATCGCCACACGCGAAGACGTATCGTGTGTCACCTTCAACTACCGTG 583
Qy	463	TTGCTCTGTGGGGTTCTTGGCTGGTATGACATCAAGCGCGAGGCGGGAACGCG 522
Dd	584	TCGCGCCCTTGGGTTCTCCTCAGCACTGGGGAC-----GCCAATCTGCCAGGTAATACTG 637
Qy	523	GCTTTGAAGCACCGCGTTTGGGCATGAGTGGGTGGCAGACAAATTCGCGGGTTCGCGG 582
Dd	638	GCCTTCGGATCAGCATGCCAATGCTTGGGTGAAGAGGAATATCGCGGCTTCGCGG 697
Qy	583	GCGACCGGAGCAAGGTGACATCTTTGGGAGGCGGGCAGCATGCC 628
Dd	698	GGGACCCCAACAACATACGCTCTTGGGGGAGTCTGTGGAGGTGC 743
RESULT 10		
US-08-445-050-1		
Sequence 1, Application US/08445050		
Patent No. 5763739		
GENERAL INFORMATION:		
APPLICANT: Blaeckberg, Lars		
APPLICANT: Edlund, Michael		
APPLICANT: Hansson, Lennart		
APPLICANT: Hernelle, Olle		
APPLICANT: Lundberg, Lennart		
APPLICANT: Stroemqvist, Mats		
APPLICANT: Toernell, Jan		
TITLE OF INVENTION: No. 5763739el Polypeptides		
NUMBER OF SEQUENCES: 21		
CORRESPONDENCE ADDRESS:		
ADDRESSEE: White & Case		
STREET: 1155 Avenue of the Americas		
CITY: New York		
STATE: New York		





Db 698 GGGACCCCAACATCAGCTTTCGGGGAGTCTGCTGGAGGTGC 743

RESULT 11  
US-08-204-691-1  
; Sequence 1, Application US/08204691  
; Patent No. 5827683  
; GENERAL INFORMATION:  
; APPLICANT: Blaeckberg, Lars  
; APPLICANT: Edlund, Michael  
; APPLICANT: Hansson, Lennart  
; APPLICANT: Hernel, Oille  
; APPLICANT: Lundberg, Lennart  
; APPLICANT: Stroemqvist, Mats  
; APPLICANT: Toernell, Jan  
; TITLE OF INVENTION: NO. 5827683el Polypeptides  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: White & Case  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10036-2787  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/204,691  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: SE 9300686-4  
; FILING DATE: 01-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: SE 9300722-7  
; FILING DATE: 04-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sterner Ph.D., Richard J  
; REGISTRATION NUMBER: 35,372  
; REFERENCE/DOCKET NUMBER: 1103326-850  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)819-8783  
; TELEFAX: (212)354-8113  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2428 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; TISSUE TYPE: mammary gland  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 82..2319  
; OTHER INFORMATION: /product= "bile-salt-stimulated  
; OTHER INFORMATION: lipase"

;  
; FEATURE: exon  
; NAME/KEY: 1576..2415  
; LOCATION:  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 151..2316  
; FEATURE:  
; NAME/KEY: polyA\_signal  
; LOCATION: 2397..2402  
; FEATURE:  
; NAME/KEY: repeat\_region  
; LOCATION: 1756..2283  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: 1..81  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 1756..1788  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 1789..1821  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 1822..1854  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 1855..1887  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 1888..1920  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 1921..1953  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 1954..1986  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 1987..2019  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 2020..2052  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 2053..2085  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 2086..2118  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 2119..2151  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 2152..2184  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 2185..2217  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 2218..2250  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 2251..2283  
; US-08-204-691-1

Query Match 4.0%; Score 61.2; DB 1; Length 2428;  
Best Local Similarity 52.7%; Pred. No. 1.3e-05;  
Matches 214; Conservative 0; Mismatches 173; Indels 19; Gaps 3;  
QY 236 TGCAGTCCAAAGGTGTTCCAGGCGGTTCCTCCAGAGTGAGGACTGCCTCACCATCAACG 295  
DB 344 TGCAGGCCACCATCACCAGGAGACACACCTACGGGGATGAGAGCTGCCTGTACCTCAACA 403

```
Qy 296 TGGTGGCGCCCGCCGCGGACCAAGCGCGCGCC---AACTCCCGGTATGCTCTGGATCT 352
Db 404 TTTGGGTGCCCGCCAGGAGAAAGTCTCCCGGACCTGCGCGTTATGATCTGGATCT 463
Qy 353 TTGGCGGTGGGTTTGAGATCGGCAGC-----CCACCATCTTCCCTCCCGCCAG 402
Db 464 ATGAGGCGCCTTCTCATGGGTCCGCGCCATGGGCGCACTTCTCAACAACCTACCTGT 523
Qy 403 ATGCTCACCAAGAGTGTGCTCATGGCAAGCATCATCCAGTGGCGGTCACTACCGTG 462
Db 524 ATGACGGCGAGGATCGCCACACCGCGAAGCTCATGCTGCTCACTTCACTTACCGTG 583
Qy 463 TTGCTCTGTGGGTTCTTGGCTGGTGATGATCAATCAAGCGCGGAGCGGGAACCGG 522
Db 584 TCGCCCGCTTGGGTTCTCTCAGCACTGGGAC-----GCCAATCTGCCAGGTAATG 637
Qy 523 GCTTGAAGGACCGCGTTTGGCATGCGGTGGTGGGAGAGCAATCTTCCGGGTTCGGG 582
Db 638 GCCTTCGGGATCAGCATGCGCCATGCTTGGGTGAAGAGGAATATCGCGCCTTCGGG 697
Qy 583 GCGACCCGAGCAAGGTGACATCTTTGGCGAGCGGCGGCGAGCATGTCC 628
Db 698 GGGACCCCAACACATCACGCTCTTCGGGGAGTCTGTGGAGTGC 743
```

## RESULT 12

```
US-09-355-295B-2
; Sequence 2, Application US/09355295B
; Patent No. 6525241
; GENERAL INFORMATION:
; APPLICANT: Dalrymple, M.
; APPLICANT: Lundberg, L.
; APPLICANT: Stromqvist, M.
; TITLE OF INVENTION: Expression Methods
; FILE REFERENCE: 1754 SEQUENCE LISTING V2A.txt
; CURRENT APPLICATION NUMBER: US/09355,295B
; PRIOR FILING DATE: 1999-07-18
; PRIOR APPLICATION NUMBER: PCT/SE99/00648
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 5
; SEQ ID NO 2
; LENGTH: 2428
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: cDNA
; FEATURE:
; OTHER INFORMATION: Mammary gland source
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 82..2319
; OTHER INFORMATION: /product= "bile salt-stimulated
; OTHER INFORMATION: lipase"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 985..1173
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1174..1377
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1378..1575
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1576..2415
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 151..2316
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 2397..2402
; FEATURE:
; NAME/KEY: repeat_region
```

```
Query Match 4.0%; Score 61.2; DB 4; Length 2428;
Best Local Similarity 52.7%; Pred. No. 1.3e-05;
Matches 214; Conservative 0; Mismatches 173; Indels 19; Gaps 3;
```

```
Qy 236 TGAGTCCAAAGGTGTTCCAGGCGGTGCTTCCCGAGAGTGAGGACTGCCTCACCATCAACG 295
Db 344 TGAGGCGCCATCACCCAGGACAGCACCTACGGGATGAAGACTGCCTGTACCTCAACA 403
Qy 296 TGGTGGCGCGCGCGGACCAAGCGCGGCGCC---AACTCCCGGTATGCTCTGGATCT 352
Db 404 TTTGGGTGCCCGCCAGGAGCAAGTCTCCCGGAGACCTGCGCGTTATGATCTGGATCT 463
Qy 353 TTGGCGGTGGGTTTGAGATCGGCAGC-----CCACCATCTTCCCTCCCGCCAG 402
Db 464 ATGAGGCGCCTTCTCATGGGTCCGCGCCATGGGCGCAACTTCTCAACAACCTACCTGT 523
Qy 403 ATGCTCACCAAGAGTGTGCTCATGGGCAAGCGCATCATCCAGTGGCGGTCAACTACCGTG 462
Db 524 ATGACGGCGAGGAGATCGCCACACAGGGAACGTATCATCGTGGTCACTTCAACTACCGTG 583
```

463 TTGCTCGTGGGGTCTTGGCTGGTGCATGACATCAAGCGCGAGCGGAGCGGGAACGCG 522  
584 TCGGCCCTTGGGTTTCTCAGCACTGGGGAC-----GCCAATCTGCCAGGTAACTATG 637  
523 GCTTGAAGGACACAGCGTTTGGGCGATGCGAGTGGGTGGCAGCAACAATTCGCCGGTTTCGGCG 582  
638 GCCTTCGGGATCAGCACATGCCAATGCTTGGGTGAAGAGGAATATCGCGCCCTTCGGGG 697  
583 GCGACCCGAGCAAGGTGACATCTTTGGCGAGCGCGGAGCATGTCC 628  
698 GGGACCCCAACAACATCACGCTCTTCGGGGAGTCTGCTGGAGGTGC 743

Db  
Qy  
Db  
Qy  
Db  
Qy  
Db

RESULT 13  
US-08-370-223-12  
; Sequence 12, Application US/08370223  
; Patent No. 6107026  
; GENERAL INFORMATION:  
; APPLICANT: Lange, III, Louis G.  
; APPLICANT: Kumar, B. Vijaya  
; TITLE OF INVENTION: Methods and Reagents for RFLP Analysis  
; TITLE OF INVENTION: of the Human Cholesterol Esterase Gene  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Allegretti & Witcoff, Ltd.  
; STREET: 10 South Wacker, Suite 3000  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/370,223  
; FILING DATE:  
; CLASSIFICATION: 307  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/053,308  
; FILING DATE:  
; APPLICATION NUMBER: 07/730,204  
; FILING DATE: July 15, 1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McDonnell, John J.  
; REGISTRATION NUMBER: 26,949  
; REFERENCE/DOCKET NUMBER: 91,441  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2487 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 104..2341  
US-08-370-223-12

Query Match 4.0%; Score 61.2; DB 3; Length 2487;  
Best Local Similarity 52.7%; Pred. No. 1.3e-05;  
Matches 214; Conservative 0; Mismatches 173; Indels 19; Gaps 3;  
236 TGCAGTCCAGGTGTTCCAGCGGTCTTCCCGAGAGTGGAGTGCCTCACCATCAACG 295  
366 TGCAGGCCACATCACCAGGACAGCACCTACGGGATGAGACTGCTGTACTCAACA 425  
296 TGGTGGCGCGCGCGGCGCACCAAGCGGGGCC---AACCTCCCGGTGATGCTGTGATCT 352

Qy  
Db  
Qy

426 TTTGGGTGCCCCAGGGGAGGAAGCAAGTCTCCCGGGACCTTCCCGTTATGATCTGATCT 485  
353 TTGGCGGTGGGTTTGAGATCGGCAGC-----CCCAACCATCTTCCCTCCCGCCCGAG 402  
486 ATGAGAGCGGCTTCTCATGGGTCCGGCCATGGGGCAACTTCTCAACAACATACCTGT 545  
403 ATGTGTACCAAGAGTGTGCTCATGGGCAAGCCATCTCCAGTGGCGGTCAACTACCGTG 462  
546 ATGACGGCGAGGAGATCGCCACACGCGGAAACGTCATCGTGGTCACTTCAACTACCGTG 605  
463 TTGCTCGTGGGGGTTCTTGGCTGGTGCATGACATCAAGCGCGAGCGGGAACGCGG 522  
606 TCGGCCCTTGGGTTCTCAGCACTGGGGAC-----GCCAATCTGCCAGGTAACTATG 659  
523 GCTTGAAGGACACAGCGTTTGGGCGATGCGAGTGGGTGGCAGCAACAATTCGCCGGTTTCGGCG 582  
660 GCCTTCGGGATCAGCACATGCCAATGCTTGGGTGAAGAGGAATATCGCGCCCTTCGGGG 719  
583 GCGACCCGAGCAAGGTGACATCTTTGGCGAGCGCGGAGCATGTCC 628  
720 GGGACCCCAACAACATCACGCTCTTCGGGGAGTCTGCTGGAGGTGC 765

Db  
Qy  
Db  
Qy  
Db  
Qy  
Db

RESULT 14  
US-07-732-962A-1  
; Sequence 1, Application US/07732962A  
; Patent No. 5248604  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Meir  
; TITLE OF INVENTION: EXPRESSION OF ENZYMATICALLY ACTIVE  
; TITLE OF INVENTION: RECOMBINANT HUMAN ACETYLCHOLINESTERASE  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John P. White, Esq.  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/732,962A  
; FILING DATE: 19910722  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 39304/JPW/LSW  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 664-0525  
; TELEX: 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1845 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1842  
US-07-732-962A-1

Query Match 3.9%; Score 60; DB 1; Length 1845;  
Best Local Similarity 53.1%; Pred. No. 2.3e-05;  
Matches 203; Conservative 0; Mismatches 165; Indels 14; Gaps 3;



Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2004, 17:10:06 ; Search time 521 Seconds  
(without alignments)  
14417.679 Million cell updates/sec

Title: US-09-943-857-3

Perfect score: 1532

Sequence: 1 gccacccgcaagctcgcca.....tgaccaacccttcttgg 1532

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 322919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1532	100.0	1532	10	US-09-943-857-3
2	1080.6	70.5	1541	10	US-09-943-857-7
3	1007.6	65.8	1511	10	US-09-943-857-9
4	998.4	65.2	1548	10	US-09-943-857-5
5	830.4	54.2	1469	10	US-09-943-857-1
6	194.8	12.7	1687	16	US-10-369-493-27923
7	88.2	5.8	1716	9	US-09-801-852A-1
8	84.6	5.5	1419	16	US-10-369-493-34501
9	82.8	5.4	1272	16	US-10-369-493-41665
10	61.2	4.0	1476	15	US-10-156-761-7341
11	61.2	4.0	2312	16	US-10-264-049-858
12	61.2	4.0	2428	9	US-09-969-347-220
13	61.2	4.0	2428	10	US-09-418-176-1
14	61.2	4.0	9025608	15	US-10-156-761-1

C	15	60	3.9	740	15	US-10-029-386-22811	Sequence 22811, A
	16	60	3.9	1725	9	US-09-810-861B-5	Sequence 5, Appli
	17	60	3.9	5767	9	US-09-810-861B-3	Sequence 3, Appli
	18	60	3.9	14446	9	US-09-810-861B-4	Sequence 4, Appli
	19	59.6	3.9	3004	10	US-09-873-367C-446	Sequence 446, App
	20	59	3.9	2220	9	US-09-801-852A-3	Sequence 3, Appli
C	21	56	3.7	1126	17	US-10-437-963-58303	Sequence 58303, A
	22	55.8	3.6	1644	13	US-10-425-114-4790	Sequence 4790, Ap
	23	55.8	3.6	3746	15	US-10-160-719-29	Sequence 29, Appl
	24	55.8	3.6	3746	15	US-10-160-719-49	Sequence 49, Appl
	25	55.8	3.6	3773	13	US-10-627-132-9	Sequence 9, Appli
	26	55.8	3.6	3773	15	US-10-209-059-9	Sequence 5, Appli
	27	55.8	3.6	3773	15	US-10-160-719-5	Sequence 3854, Ap
	28	53.4	3.5	2191	9	US-09-880-107-3854	Sequence 1038, Ap
	29	53.4	3.5	2191	10	US-09-873-367C-828	Sequence 828, App
	30	53.4	3.5	2191	10	US-09-873-367C-829	Sequence 829, App
	31	53.4	3.5	2484	13	US-09-925-298-271	Sequence 271, App
	32	53.4	3.5	2484	15	US-10-102-806-271	Sequence 271, App
	33	53.4	3.5	2484	15	US-10-369-493-40512	Sequence 40512, A
	34	52.2	3.4	1548	16	US-10-104-047-249	Sequence 249, App
	35	52	3.4	2092	16	US-10-437-963-32084	Sequence 32084, A
	36	51.4	3.4	1306	17	US-10-451-168-40	Sequence 40, Appl
	37	51.2	3.3	1728	17	US-10-114-270-195	Sequence 195, App
	38	51.2	3.3	1746	13	US-10-023-515-3	Sequence 3, Appli
	39	51.2	3.3	1746	14	US-10-023-515-3	Sequence 3, Appli
	40	51.2	3.3	1746	17	US-10-674-636-3	Sequence 3, Appli
	41	51.2	3.3	1746	17	US-10-451-168-41	Sequence 41, Appl
	42	51.2	3.3	1857	17	US-10-381-898-20	Sequence 20, Appl
	43	51.2	3.3	1962	16	US-10-094-749-736	Sequence 736, App
	44	51.2	3.3	2158	14	US-10-023-515-1	Sequence 1, Appli
	45	51.2	3.3	2158	17	US-10-674-636-1	Sequence 1, Appli

#### ALIGNMENTS

#### RESULT 1

US-09-943-857-3  
; Sequence 3, Application US/09943857  
; Publication No. US20030124701A1  
; GENERAL INFORMATION:  
; APPLICANT: Shaw, Jai-Fu  
; APPLICANT: Lee, Guan-Chiun  
; APPLICANT: Tang, Shye-Jye  
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES  
; FILE REFERENCE: 08919-066001  
; CURRENT APPLICATION NUMBER: US/09/943,857  
; CURRENT FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1532  
; TYPE: DNA  
; ORGANISM: Candida rugosa  
US-09-943-857-3

Query Match 100.0%; Score 1532; DB 10; Length 1532;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GCCACCGCAAGCTCGCCAAACGGCGACACCATCACCGGTCTCAACGGCCATCATCAACGA	60
DB	1	GCCACCGCAAGCTCGCCAAACGGCGACACCATCACCGGTCTCAACGGCCATCATCAACGA	60
QY	61	GGCGTTCCTCGCATTCCTTTGCCAGCGCGGTGGCAACCTTCGGTTCAAGSACCC	120
DB	61	GGCGTTCCTCGCATTCCTTTGCCAGCGCGGTGGCAACCTTCGGTTCAAGSACCC	120
QY	121	TGTGCGTACTCTGCTGCTCAAGCGGCAGAACTTACTAGCGCGGTGATGAGCAG	180
DB	121	TGTGCGTACTCTGCTGCTCAAGCGGCAGAACTTACTAGCGCGGTGATGAGCAG	180
QY	181	AACCCCGAGGGCAAGTTTGAAGAGAACCTTGGCAAGACGGCACTCGACTTGGTGTGATGAG	240

```
Db 181 AACCCGAGGCGAGTTTGAAGAACTTTGGCAAGCGCACTCGACTGGTGTATGAG 240
Qy 241 TCCAAAGTGTTCAGGCGGTGCTTCCCAAGAGTAGGACTGCTCACCATCAACAGTGGT 300
Db 241 TCCAAAGTGTTCAGGCGGTGCTTCCCAAGAGTAGGACTGCTCACCATCAACAGTGGT 300
Qy 301 CGGCGCGCGGCGACCAAGGCGGCGCAACCTCCCGGTTCATGCTCTGGATCTTTGGCGGT 360
Db 301 CGGCGCGCGGCGACCAAGGCGGCGCAACCTCCCGGTTCATGCTCTGGATCTTTGGCGGT 360
Qy 361 GGGTTTGAGATCGGAGCGCCACCATCTTCCCTCCCGCCAGATGGTCAACAAGTGTG 420
Db 361 GGGTTTGAGATCGGAGCGCCACCATCTTCCCTCCCGCCAGATGGTCAACAAGTGTG 420
Qy 421 CTATGGGCAAGCCATCATCATGCTGGCGGTCAACTACCGTGTTCCTCGTGGGGTCT 480
Db 421 CTATGGGCAAGCCATCATCATGCTGGCGGTCAACTACCGTGTTCCTCGTGGGGTCT 480
Qy 481 TGGCTGTGTGATGATCAAGGCCGAGGCGAGCGGAAACCGCGCTTTGAAGACACAGCGTT 540
Db 481 TGGCTGTGTGATGATCAAGGCCGAGGCGAGCGGAAACCGCGCTTTGAAGACACAGCGTT 540
Qy 541 TGGGATGTCAGTGGGTGGCAGACAATTCGCGGTTTCGGCGGACCCGAGCAAGTGA 600
Db 541 TGGGATGTCAGTGGGTGGCAGACAATTCGCGGTTTCGGCGGACCCGAGCAAGTGA 600
Qy 601 CATCTTTGGGAGGCGGCGAGTGTGCGTGTTCGCGGCTCATCTCATCTGGAACGAGCGGA 660
Db 601 CATCTTTGGGAGGCGGCGAGTGTGCGTGTTCGCGGCTCATCTCATCTGGAACGAGCGGA 660
Qy 661 CAACAGTACAGGCAAGCGTGTTCGCGGCGGCGATCATGAGGAGCGCATGGTGCC 720
Db 661 CAACAGTACAGGCAAGCGTGTTCGCGGCGGCGATCATGAGGAGCGCATGGTGCC 720
Qy 721 GGACCCGTTGAGCGGACGTACGGCAACAGATCTACGACCTTTTGTCTCGAGTGTGG 780
Db 721 GGACCCGTTGAGCGGACGTACGGCAACAGATCTACGACCTTTTGTCTCGAGTGTGG 780
Qy 781 CTGTGGAGCGGCGAGCAAGCTCGGTGCTTGGGAGTGGGAGCGACACTTGTCTGA 840
Db 781 CTGTGGAGCGGCGAGCAAGCTCGGTGCTTGGGAGTGGGAGCGACACTTGTCTGA 840
Qy 841 TGCCACCAACAACTCTCTGGTCTTCGCGTACTCTCTGCTTGGCTGTACTCCCGGCC 900
Db 841 TGCCACCAACAACTCTCTGGTCTTCGCGTACTCTCTGCTTGGCTGTACTCCCGGCC 900
Qy 901 CGACGGCAAGAACATCACCGATGACATGTACAAGTTGGTGGCGGACGGCAAGTATGGAAG 960
Db 901 CGACGGCAAGAACATCACCGATGACATGTACAAGTTGGTGGCGGACGGCAAGTATGGAAG 960
Qy 961 GGTTCCTGATCATTTGGGACGAGACGAGGCGGACCACTTTGGCTCTTGAACGTG 1020
Db 961 GGTTCCTGATCATTTGGGACGAGACGAGGCGGACCACTTTGGCTCTTGAACGTG 1020
Qy 1021 ACCAGATGTCTAGGCGGCTGCTTACTTCAAGCAGTTCATCCAGCCGAGCGAGCGGAG 1080
Db 1021 ACCAGATGTCTAGGCGGCTGCTTACTTCAAGCAGTTCATCCAGCGGACGCGGAG 1080
Qy 1081 ATCGACACTTGTAGGCGGCTGACCCCAAGGACATCAACCGAGGTCGTTGACACGGTT 1140
Db 1081 ATCGACACTTGTAGGCGGCTGACCCCAAGGACATCAACCGAGGTCGTTGACACGGTT 1140
Qy 1141 CAAAGCTCACCCGCGAGTCAAGAGATCGCGGTGCTCGGCGACCTTGCATTCATCCACG 1200
Db 1141 CAAAGCTCACCCGCGAGTCAAGAGATCGCGGTGCTCGGCGACCTTGCATTCATCCACG 1200
Qy 1201 CCGCGCGTACTTCTCAACCACTTCCAGGCGGCGACCAAGTACTCGTTCTTCAAGCAGC 1260
Db 1201 CCGCGCGTACTTCTCAACCACTTCCAGGCGGCGACCAAGTACTCGTTCTTCAAGCAGC 1260
Qy 1261 TCGGGTGCATCATGAGGCGACCTTCCATGCGCAAGACATGTTGTGGGAGGACTACTTGT 1320
```

```
Db 1261 TCGGGTGCATCATGAGGCGACCTTCCATGCCAAGCAGATTTGTGTGGAGGACTACTTGT 1320
Qy 1321 TGGGAAGCGGCGAGCGTCACTTACAAACAAACGCGTTTATCGGCTTCGCCACCGACTTGGACC 1380
Db 1321 TGGGAAGCGGCGAGCGTCACTTACAAACAAACGCGTTTATCGGCTTCGCCACCGACTTGGACC 1380
Qy 1381 CCAACACCGCGGGGTTGTTGGTGAATGCGGCCAAGTACACGAGCAGCCAGGGCAACAAC 1440
Db 1381 CCAACACCGCGGGGTTGTTGGTGAATGCGGCCAAGTACACGAGCAGCCAGGGCAACAAC 1440
Qy 1441 TGATGATGATCAACGCGCTTGGGCTTGACACCGGCAAGGCAACCTTCGCAACCGCTGGCT 1500
Db 1441 TGATGATGATCAACGCGCTTGGGCTTGACACCGGCAAGGCAACCTTCGCAACCGCTGGCT 1500
Qy 1501 ACGACGCGGTGATGACCAACCCGTTCTTTGTG 1532
Db 1501 ACGACGCGGTGATGACCAACCCGTTCTTTGTG 1532

RESULT 2
US-09-943-857-7
; Sequence 7, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jei-Pu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1541
; TYPE: DNA
; ORGANISM: Candida rugosa
US-09-943-857-7

Query Match 70.5%; Score 1080.6; DB 10; Length 1541;
Best Local Similarity 84.9%; Pred. No. 7.9e-294;
Matches 1311; Conservative 0; Mismatches 209; Indels 25; Gaps 8;
```

```
Qy 3 CCACCGCCAAAGCTCGCCAAACGGCGGACACCATACCGGTCTCAACGCCATCATCAACGAGG 62
Db 2 CCCAGCCACGCTCGCCAAACGGCGGACACCATACCGGTCTCAACGCCATCATCAACGAGG 61
Qy 63 GTTCTCTGGCATTCCTTTGCGGAGCGCGGTGGGCAACCTCCGCTTCAAGGACCCCTG 122
Db 62 GTTCTCTGGCATTCCTTTGCGGAGCGCGGTGGGCAACCTCCGCTTCAAGGACCCCTG 121
Qy 123 TGCCTACTCTGGCTCGCTCAACGGCCAGAAAGTT----ACTTACGGCCCGCTGCATGCAGC 178
Db 122 TGCCTACTCTGGGTCTCTCAACGGTCAATCTTACCGCGTACGGTCCGTCATGCAGC 181
Qy 179 AGAACCCCGAGGCGACGTTTGAAGAGAACCTTGGCAAGACGGGACATCGACTTGGTATGC 238
Db 182 AGAACCCCGAGGCGACCTACGAGGAGAACCTCCCAAGTGGCGCTTGAATTTGGTATGC 241
Qy 239 AGTCCAAGGTGTTCCAGGCGGTGTTCCCGAGAGTGGAGTCCCTCAACCAACGTTGG 298
Db 242 AGTCCAAGGTGTTCCAGGCTGTTCTTCCCAACAGGAGGAGTCCCTCAACCAACGTTGG 301
Qy 299 TCGCGCCCGCGGCGACCAAGCGGCGCGCAACCTCCCGGTCACTGTCTGGATCTTTGGCG 358
Db 302 TCGCGCCCGCGGCGACCAAGCGGCGCGCAACCTCCCGGTCACTGTCTGGATCTTTGGCG 361
Qy 359 GTGGGTTTGAAGATCGGAGCGCCCAACATCTTCCCTCCCGCCAGATGTCACCAAGAGTG 418
Db 362 GTGGGTTTGAAGATCGGAGCGCCCAACATCTTCCCTCCCGCTCAGATGTTCTCCAAGAGTG 421
Qy 419 TGCTCATGGGCAAGGCCATCATCCAGTGGCGGCTCAACTACCGTGTTCCTCGTGGGGT 478
```





```

QY 721 GGACCCGGTGGACGGACAGTACGGCAACAGATCTACAGCTCTTTGTCGAGTCTGG 780
Db 715 GGACGC-GTGGACGGCTC-TACGGCAACAGATCTTTGACCTCTTGGCTGCA-CGCGGG 771
QY 781 CTGTGGCAGCGCAGACGACAGCTCGCGTGTCTCGCAGTGGCAGCGACACCTTGTCTCGA 840
Db 772 CTGCGGACGGCCAGCAGCAAGCTTGTGCTTGGCGGTGTGAGCGACAGTGTGAGGA 831
QY 841 TGCCACCAACAACACTCTCGGTCTTGGCGTACTCTCTCGTGTGGTGTACT-CCGCGC 899
Db 832 CGCACCAACAACACCCCTGGTCTTGGCGTACTCTCTCGTGTGGTGTACTCTCGCGC 891
QY 900 CGGACGCAAGAACATCACCGATGACATGTACAGTGTGTGGCGGCGGCAAGTATCAA 959
Db 892 CCGACGGCGTGAACATCACCGACACATGTGC---CTTGTGCGGCGGCAAGTATGCA- 947
QY 960 GCGTTCCTCGTATCATTTGGGACAGACAGAGGCGACCATCTTTGGC-TCTTGAACG 1018
Db 948 ---CTCTGTGATCATCGCGACCCAGAACGAGAGGCGACCTTCTTTGGCACCTTGAACG 1004
QY 1019 TGACCAAGATGTCTAGCGCCCGTGTACTTCAAGAGTTCATCCAGCGCAGCGACGGG 1078
Db 1005 TGACCAAGATGTCTAGCGCCCGTGTACTTCAAGAGTTCATCCAGCGCAGCGACGGG 1061
QY 1079 AGATCGACACCTTGATGGCGGCTACCCCGAGGACATCACCGAGGCTCGGTTCGACACGG 1138
Db 1062 AG-TGACACAGTGTGATGACGGGTACCCCGGACATCACCGAGGCGGTTTCGACACGGT 1120
QY 1139 TTCAACG-CTCACCCCGAGTCAAGAGATCGCGTGTCTCGGCGACCTTGCATTCATC 1196
Db 1121 CTCACGCGCTCACCCCGAGTCAAGAGATCGCGTGTCTCGGCGACCTTGTCTAAG- 1178
QY 1197 CACGCGCCCGCTACTCTCAACCACTTCCAGGGCGGCAAGTACTCTGTCTCAAG 1256
Db 1179 -----CCGTGCTACTCTCAACCACTTCCAGGGCGGCAAGTACTCTGTCTCAAG 1233
QY 1257 CAGCTCGGGTGGCAATCATGGGACCTTCCATGCGCAAGCATTTGTGCGAGACTAC 1316
Db 1234 CAGCTCGGGTGGCGGGTCTCGGAACGTTTCACTCAACAGCATTTGTCTTCAGAGCTAC 1293
QY 1317 TTGTGGGAAGCGGCGAGCGTCTATCTACAAACAGCGTTTATCGCGTTCGCCACCGATTG 1376
Db 1294 TTGTGGGACGGCTCGCTCATCTACAAACAGCGTTTATCGCGTTCGCCACCGATTG 1353
QY 1377 GACCCCAACACCGCGGGTGTGTGTGAACCTGGGCCCAAGTACACAGCAGCGCGGCAAC 1436
Db 1354 GACCCCAACACCGCGGGTGTGTGTGAAGTGGCCCGAGTACACAGCAGCGCGGCGCAAC 1413
QY 1437 AACTTGATGATGATCAACGCTTGGCTGTGTACCGCGCAAGCAACTTTCGACCGCT 1496
Db 1414 AACTTGATGATGATCAACGCTTGGCTGTGTACCGCGGAGTGTACCGCGGAGGACAC- 1472
QY 1497 GGCTACGAGCGGTTGATGACCAACCGCTTCTT 1528
Db 1473 GGCTACGAGCGGTTTCTTCAACCGCGGTT 1504

```

RESULT 4

```

US-09-943-857-5
; Sequence 5, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jui-Fu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5

```

```

; LENGTH: 1548
; TYPE: DNA
; ORGANISM: Candida rugosa
US-09-943-857-5
Query Match 65.2%; Score 998.4; DB 10; Length 1548;
Best Local Similarity 83.8%; Pred. No. 1.1e-270;
Matches 1301; Conservative 0; Mismatches 221; Indels 30; Gaps 14;

QY 1 GCCACCGCCAAAGCTCGCCAAACGGGACACCATCACCGGTCTCAACGCCATCATCAACGA 60
Db 1 GCCACATGCCACGCTCGCCAAACGGGACACCATCACCGGTCTCAACGCCATCATCAACGA 60
QY 61 GGGGTTCTCTCGGCATTTCCCTTTGCGGAGCCGCGGTGGGCAACCTCCGCTTCAAGAGCCC 120
Db 61 GGGGTTCTCTCGGTATTTCCCTTTGCTCAGCCGCGGTGGGCAACCTCCGCTTCAAGAGCCC 120
QY 121 TGTGCGGTACTTGTGGTCTGCTCAACGGCGAGAAGTT---ACTTACGGCCCGTGTGATCGAG 177
Db 121 TGTGCGGTACTTGTGGTCTGCTCAATGCTGAGAAGTTTACTGTATGGCCCTGTGATCGAG 180
QY 178 CAGAACCCCGAGGCGACGTTTGAAGAAACCTTGGCAAGACGGCACTCGACTTGGTGTATG 237
Db 181 ATGAACCAATTTGGCAACTGGGACTCTCTGCTTCCCHAGGCTGCAATCAACT---TGATG 237
QY 238 CAGTCCAAAGTGTTCAGGGGGTGTCTTCCCAAGAGTGAAGAGTGAAGTCTCCATCAACG 297
Db 238 CAGTCCAAAGTGTTCAGGGGGTGTCTTCCCAAGAGTGAAGAGTGAAGTCTCCATCAACG 297
QY 298 GTGCGGCGCGGCGACCAAGCGGGCGGCAACCTCCCGGTCTGCTGTGATCTTTGGC 357
Db 298 GTGCG---GCCGGGCGACCAAGCGGGTGCACCACTCCCGGTGTGATGTTGGTATTTTGGC 354
QY 358 GGTGGGTTTGAAGTTCGGCAGCGCCCAACCTCTTCCCTCCCGCCAGATGGTCAACAGAGT 417
Db 355 GCGCGGTTTGAAGTTCGGGCTCCAGTCTCTTCCCTCCCGCACAGATGATCAACGCCAGC 414
QY 418 GTGCTCATGGCAAG-CCATCATCAAGTGGCGGTAACTACCGTGTTCCTTCGTGGGG 476
Db 415 GTGCTTATGGCAAGGCCCATCATCAAGTGAAGTGAAGTGAAGTGTTCGTGGGG 474
QY 477 TCTTGGTGTGATGATCAAGCGGCGGCGGCAACCGCGGTCTCAAGGACCGAG 536
Db 475 TCTTGGTGTGATGATCAAGCGGCGGCGGCAACCGCGGTGTGCAACCGAG 534
QY 537 GGTGGGCGATGAGTGGTGGCAGCAAACTTTCGGCGGTTCGGCGGCGACCCGAGCAAG 596
Db 535 CGCTTGGGTTTTCAGTGGGTGGCGCAAACTTTCGGCGGTTCGGCGGCGACCCGAGCAAG 594
QY 597 GTGA-CATCTTTTGGCGA-GGCGGGCGAGCATGTCCGCTGTGTGCGCACTCATCTGGAACGA 654
Db 595 GTGACCATCTTTTGGTGGGCGGCGAGCATGTTCGGTAAATGTGTCAAGTCTCTCTGGAACGA 654
QY 655 CGGCGCAACAGCTTACAGGCGAGCGGCGGTTCGGCGGCGGCACTCATGAGGAGCGCAT 714
Db 655 CGGCGCAACAGCTTACAGGCGAGCGGCGGTTCGGCGGCGGCACTCATGAGGAGCGCAT 714
QY 715 GGTGCG--GGACCCCGGTGGACCGGCACTGTACGGCAACAGAGATCTACGACTCTTTGTCTCA 773
Db 715 GGTGCGGGGACCCCGGTGGATGGGCGCTTACGGCGACAGATCTACGAGGTGTGTGCTT 774
QY 774 GTGCTGGTGTGGGCGAGCGCGAGCGCAAGCTCGCGTGTGTGCGGAGTG-CGAGCGACACC 832
Db 775 CAGCGCGGTGTGGCAGTGCAGCGCAAGCTCGCGTGTGTGCGGAGTGTGCGGAGTGTGCGG 834
QY 833 TTGCTCGATGCCACCAACAACACTCTCTGGTCTTTGGCGTACTCTCTGCTTGGCGGTG--- 889
Db 835 CTCCTTCAGGCGCACCGAGCGACTCTCGGGGGCTTGGCGGCTACCTCTCTGCTGCGGTGTT 894
QY 890 -TACTCCCGCGCGCGGCGGCAAGCAATCACCGATGACATGTATCAAGTGTGTGCGCGACGG 948
Db 895 TCTCCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 954

```

```
QY 949 CAAGTATGCAAGCTTCCCGTGATCATTTGGCGAACAGAGGACCACTTTTG- 1007
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
955 CAAGTGTGCCAACCTTTCCGGTGATCATTTGGCGAACAGAGGACCACTTTTG 1014
QY 1008 GCTCTCAAGCTGACCAAGATGCTCAGGCGGTGCTTACTTCAAGCAGTTTCAACGC 1067
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1015 GTTGTGAAAGTGACTACGGATGCTCAGGACGCGGAGTACTTCAAGGAATTCATCCAGC 1074
QY 1068 CAGCGACGCGAGATGACACATTTGATGGCGGCGTACCCCGAGGACATCAACAGGGTCC 1127
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1075 CAGCGACGCGAGATGACACATTTGATGGCGGCGTACCCCGAGGACATCAACAGGGTCC 1134
QY 1128 GTTGCACACGG-----TTCAACGC--TCACCCCGAGTTCAGAGAAT---CGCGGTGCT 1177
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1135 GTTGCACACCGGATCTTCAACGCATCAACCGCGAGTTCAAACGGATTCAGCGGTGCT 1194
QY 1178 CGCGACCTTGCATTCATCAACGCGCGCGGTACTCTCAACCACTTCCAGGCGGAC 1237
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1195 TGGTGACCTTGGTTACTCTCCCCCGGCGTACTTCTCAACCACTTCCAGGCGGAC 1254
QY 1238 CAAGTACTCGTTCTC--AAGCAGCTCGGGTTGCCAAATCATGGGCACTTCCATGCCAAAG 1296
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1255 CAAGTACTCGTTCTCGAAGCAGCTTGGTTGCCGCTGATGGACCCACCAAG 1314
QY 1297 ACATTGTGGAGGACTACTTGTGGGAAGCGGACGCTCATCTACAACACGCTTTA 1356
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1315 ACATTGTGGAGGACTTTTGTGAGCCACAGCAGCGCGTGTACAACAACGCTTTA 1374
QY 1357 TCGGTTTCGACCGACTTCGACCCCAACCGCGGGTGTGGTGAACCTGGCGCAAGT 1416
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1375 TTGCTTTGCCAAGCAGCTCGACCCGAAAGCGCGGTGTGTGAACCTGGCGCAAGT 1434
QY 1417 ACACGACGACCGAGGCAACAATCTGATGATGATCAACGCTTGGGCTGTACACCGCA 1476
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1435 ACACGACGACCGAGGCAACAATCTGATGATGATCAACGCTTGGGCTGTACACCGCA 1494
QY 1477 AGACAACTTCGACCGCTGGCTACGACCGGTGTGACCAACCGCTTCT 1528
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1495 AGACAACTTCGACCGCTGGCTACGACCGGTGTGTTACCAACCGCTTTT 1546

RESULT 5
US-09-943-857-1
; Sequence 1, Application US/09943857
; Publication No. US20030124701A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Jai-Fu
; APPLICANT: Lee, Guan-Chiun
; APPLICANT: Tang, Shye-Jye
; TITLE OF INVENTION: RECOMBINANT CANDIDA RUGOSA LIPASES
; FILE REFERENCE: 08919-066001
; CURRENT APPLICATION NUMBER: US/09/943,857
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1469
; TYPE: DNA
; ORGANISM: Candida rugosa
US-09-943-857-1
```

```
Query Match 54.2%; Score 830.4; DB 10; Length 1469;
Best Local Similarity 78.1%; Pred. No. 2.1e-223;
Matches 1202; Conservative 0; Mismatches 241; Indels 97; Gaps 13;

QY 10 CAAGTTCGCCAACGGGACACCATACCGGTCTCAACGGCATCATCAACGAGCGTTCCT 69
Db 8 CCACCTCGCCAAAGCGGACACCATACCGGTCTCAACGGCATGTCAACGAAAAGTTTCT 67

QY 70 CGGCATTCCTTTGCCAGCGCGGTGGGCAACCTCCGCTTCAAGGACCTGTGCGGTA 129
Db 68 CGGCATTCCTTTGCCAGCGCGGTGGGCA--CTCCGCTTCAA----- 110
```

```
QY 130 CTCTGGCTCGCTCAACGCGCCAGAAG--TTACTTACGGCCCGTGATCGACGAGAACCCCGA 188
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
111 -----GCTCAACGGCCAGAGTTTACTACGCGCCGTCATGCAATGAACCTTAT 161
QY 189 GGGCACGCTTTGAAGAGAACCTTTGGCAAGACGGCACTCGACTTTGGTGTGATGCAAGT 248
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
162 GGGCTCGTTTCA-----TTGGTGTCTCCAGTCCAAGAT 193
QY 249 GTTCCAGCGGTGTTCCCGCAGAGTGAAGACTGCTCTACCATCAACGTGTGCGCGCCGCC 308
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
194 CTTTCAAGTGTGTTCTCCCAACGACGAGGACTGTCTCAC----- 232
QY 309 GGGCACCAAGCGCGCGCCCAACCTCCCGGCTCATCTCTGGATCTTTGCGCGTGGTTGA 368
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
233 ---CACCAGGGCCAGTGTCTGCTCTCCCGGTGATGCTCTGGATCTTTGGCGGTGGTTGA 289
QY 369 GATCGGAGCGCCACCACTTTCCTCCCGCCAGAGTGTCAACAGAGTGTGCTCATGGG 428
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
290 GCTTGGCGGTCTCAGCCTCTTTCAGGAGACAGATGTTGGCCAAAGAGCGTGTCTATGGG 349
QY 429 CAAGCC--ATCATCCAGTGGCGCTCAACTACCGTGTGCTCTGTTGGGGGTCTTTGGCTGG 487
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
350 TAAACCGGTGATCCAGTGTGAGCATGAATACCGCGTGGCGTCAATGGGGGTCTTTGGCGGG 409
QY 488 TGATGACATCAAGCGCGGAGGCGAGCGGAAACGCGGCTTGAAGACCAAGAGTGTGGGCAT 547
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
410 CCCCAGATCCAGAAAGAGGAGCGGAAACGCGGCTTGCATGACCAAGCGCTTGGCCAT 469
QY 548 GCGTGGGTGGGAGACAAATTCGCGGTTTCGCGGCGACCCGAGCAAGGTGA--CATCTT 606
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
470 GCAGTGGGTGGCGAGCAAACTTCTGGGTTTGGCGGCGACCCGAGCAAGGTGACCATATA 529
QY 607 TGGCGAGGCGGCGAGCATGTCTCGTGTGTCGACCTCATCTGAAACGACGCGCAACAC 666
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
530 CGCGAGGCGGCGAGCATGTGCGACGTTTGTGCACTTGTGTGAAACGACGCGCAACAC 589
QY 667 GTAAAGGGCAAGCGCTTGTTCGCGCGGCGCATCATGCGAGGAGCCATGTGCGGACCC 726
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
590 GTAAAGGGCAAGCGCTTGTTCGCGCGGCGCATCATGCGAGGCTGCAATGTGCGGACCC 649
QY 727 GTTGAACGCGACGTACCGCAACGAGATCTACGACCTCTTTTGTCTCGAGTCTGCTGTGG 786
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
650 GGTGGAACGCGACGTACCGCAACGAGATCTACAAACGAGTGTGCGCTGTCGCGGTGG 709
QY 787 CAGCGCGAGGACAAAGCTGCGTGTGTCGCGAG---TGGAGCGACACTTGTCTCGATGC 843
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
710 CAGTGCAGCGACAAGCTGCGTGTGCGCGGCTTCTCTGAGGACACGTTGTACCGGC 769
QY 844 CACCAACAACTCTCTCGGTTCCTTGGCTACTCTCTGTTGCGTGTGA---CTCCCGGCC 900
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
770 CAGGAGCGACACGCGCGGCTGTTGGCGTACCGCTGTTGCGGTGTTGTTCTCCCGCGCC 829
QY 901 CGACGCAAGAAACATCACCGATGACATGTACAAGTTTGGTGGCGACGCGCAAGTATCAAG 960
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
830 CGACGCACTTCTCATCACGACGACATGTATGCTTGGTGGCGACGCGCAAGTATCAAGCA 889
QY 961 CGTTCCTGATCATTGGCGACGAGAACGACGAGGACCATCTTT--GGCTTTGAAGCT 1019
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
890 CGTGGCGGTGATCATCGCGGACGAGAACGACGAGGCGCACTTTTGTGGGCTCTTGAAGCT 949
QY 1020 GACCAAGATGCTCAGGCGCGTCTTACTTCAAGCAGTTTCAACGCGGACGCGCA 1079
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
950 GACCAAGATGCTCAGGCGCGGTACTTCAAGCAGTTTCAACGCGGAGTCCGGA 1009
QY 1080 GATCGACACTTGTATGGCGCGTACCCCGAGGACATCACCCAGGCTTCCGTCGACACG-- 1138
Db 1010 GATCGACACTTGTATGGCGCGTACACGACGACATCACCCAGGCTTCCGTCGACACGCG 1069
QY 1139 -----TTCAACGCTCACCCCGAGTTCAGAGAAATCGCGGTGCTCGGCGACTTGCATT 1192
Db 1070 CATCTTCAATGCATCACCCCGAGTTCAAACGAGTTCGCGTGTGTTGGCGACTTGCCT 1129
QY 1193 CATCCAGCGCGCGCTACTTCTCTCAACCACTTCCAGGCGCGCACCAAGTACTCGTCTCT 1252
```

Db 1130 CAGCTTGGCGTCCGCTCTCTCTCACTACTACAGGGCGCACCAAGTACTCGTT-CT 1188  
Qy 1253 CAAGCAGCTCGGTTGCAATCATATGCGCACTTCCATGCGCAACAGCATTTGTGGCAGGA 1312  
Db 1189 CAAGCAGCTTGGTTGGCGCTTGGCGCACTTCCAGCGCAACGACATCATCTGGCAGGA 1248  
Qy 1313 CTACTTGTGGGAAGCGGCGAGCTCATCAACAACGCGTTTATCGGTTTCGCCACCGA 1372  
Db 1249 CTACTTGTGGGCGAGCGGCGAGTGATCTACAACAACGCGTTTATCGGTTTGCACGA 1308  
Qy 1373 CTTCGACCCCAACACCGCGGCTTGTGGTCAACTGGCCCAAGTACACAGCAGCGAGG 1432  
Db 1309 CCTCGACCCGCAACAGCGGCTTGTGACCACTGGCCCAAGTACACAGCAG-CAGGG 1367  
Qy 1433 CAACAACCTTGATGATGATCAACGCGCTTGGCTTGTACACCGGCAAGGACAACTTCCGCAC 1492  
Db 1368 CAACAACCTTGATGATGATCAACGCGCTTGGGCTTGTACACCGGCAAGGACAACTTCCGCC 1427  
Qy 1493 CGCTGGCTACAGCGGTTGATGACCAACCGCTTCTTTGTG 1532  
Db 1428 GGATGCGTACAGCGGCTTCTTTTCAACCGGCTTCTTTG 1467

RESULT 6

US-10-369-493-27923  
; Sequence 27923, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 27923  
; LENGTH: 1687  
; TYPE: DNA  
; ORGANISM: Neurospora crassa  
US-10-369-493-27923

Query Match 12.7%; Score 194.8; DB 16; Length 1687;  
Best Local Similarity 55.6%; Pred. No. 2e-44;  
Matches 589; Conservative 0; Mismatches 392; Indels 79; Gaps 8;  
Qy 25 CGACACCATCACCGGTTCTCAAGCCATCATACAGAGGGCTTCTCGGATTCCTTTCG 84  
Db 27 CGGTACGATTTGGTGTCAACGCGCATCTCTCACTAGAGGCTTCAACGGAATCCCTACGC 86  
Qy 85 CGAGCCGCGGTGGCAACTCGCTTCAAGGACCTGTGCGGTCTCTCTGCTCGCTCAA 144  
Db 87 CCTTCTCCGACCGCAACCTTCGCTCAAGCTCCCGTGAGACTTAAGTGTCTCTGG 146  
Qy 145 CGGCCAGAGTTACTTA-----CGGCCGCTGCATGCAGCAGAACCCCGAGGCGCAGTTGA 200  
Db 147 TGTCTTTGATGCTGTGGCATCGGCTTGTGCTCCCGGCTTCTTGTGACACCTCGTC 206  
Qy 201 AGAAGACCTTTGGCAAGACGCACTCGACTTGTGTGATGAGTCAAGGTGTTCAGGCGGT 260  
Db 207 GAACGAGTTTCTGCTCAGGTTATCGATAAGATCGTTAAACACGAGCTTTTCAAGACTAT 266  
Qy 261 GCTTCCCGAGTGAAGTACGCTTCAACCATCAAGTGTGCGCGCGCGGCGCAAGGC 320  
Db 267 ACTCAAGCTCAAGAGACTGCTTGAACCATCTCGGTCACTCGTCCCAAGGCGCAAGGC 326  
Qy 321 GGGCGCCAACTCTCCGCTCATGCTCTGATCTTTGGCGGTGGGTTTG----- 367

Db 327 TGGTGAATAGCTCCCGTCTCTTTCTGATCTTTTGGTGGTGGTTTGAAGTGAATAATCC 386  
Qy 368 -----AGATCGGACGCCCAACCA 385  
Db 387 AGCTTATATAGCGATGTAATGAACAGTGTCTAAACATTCACAGCTCGGATCGGCTCCA 446  
Qy 386 TCTTCCCTCCCGCCAGATGTCACCAAGAGTGTCTCATGGGCAAGCCAT-CATCCACG 444  
Db 447 TGTACGATGCGCTCCCTAGTCACCAACGCTATCAACAGGTAAGCCGTACGCTTACG 506  
Qy 445 TGGCGCTCAACTACCGTGTGCTCGTGGGGTCTTGGCTGTGTGATGATCAATCAAGGCGG 504  
Db 507 TTGCGGTCAACTACCGTGTGCTCGTGGTTCATGCCGGAAGAGATCTTAAAG 566  
Qy 505 AGGCGAGCGGCAACCGCGCTTGAAGGACCAAGCGTTTGGGCATGCAAGTGTGGGCGG 564  
Db 567 ACGGCTTCTCAACTTGGGTCACTTGAACGCGCATGGGCTCCAGTGGGTGGCGACA 626  
Qy 565 ACATTGGCGGTTTCGGCGGCGACCCGAGCAAGGTGAC-ATCTTTGGCGAG--CGGGCA 620  
Db 627 ACATTGCTGCTTCCGCGGTGACCCAGCAAGGTCACTATCTGGGGCGAGTCCGCGGTG 686  
Qy 621 GCATGTCGCTGTGTGCGACCTCATCTGGAACGAGCGGCAACAACATGAGGCGCAAGC 680  
Db 687 CCATGTCGCTTCAACCGAGTGTCTCTATGACGCTGACCAACAGTACAAAGCGCAAGC 746  
Qy 681 CGTTGTTCCGCGCGGCTCATGCGAGGAG-----CCATGTCGCGGACCCCGGTGGAGG 734  
Db 747 CCCTTTCCGTGGCGCCATCATGAATCTGCTTCCATCGTCCCGCGCGCTCGACT 806  
Qy 735 GCAGTACGCGCAACAGATCTACGACCTTCTGTCTCAAGTGTCTGCTGTGGCAGCGCA 794  
Db 807 GCCCAAGGCGCAGAAAGTCTACGACACCGCTGCTCAAGAACGCGCGGTCTCTGTGTG 866  
Qy 795 GCGCAAGCTCGCGTGTGGC-----AGTGGAGCGACACCTTGTCTCATGCCACCAACA 851  
Db 867 CTGACACCTTGTGCTTGGCTGCGGCTTCTTCCCTAGACACTTCTTCAAGGCGCTACT 926  
Qy 852 ACACCTCTGCGGTTCTTGGCGTACTCTCTGTTGCGGTTGT-----ACTCCCGCGCGGAGG 905  
Db 927 CGTGCCTGGGATCTGCTGACAACTCCGTGCTCTTTTACCTCCCGCGACCGGATG 986  
Qy 906 GCAAGAACATCACCGATGACATGATCAAGTGTGGTGGCGACGCGCAAGTATCAAGCGTTC 965  
Db 987 GCAAGGCTTGACTCAGAGCGCGGATAAGCTCATGCTCGCTAAGAGTACGCGCGCTCC 1046  
Qy 966 CGGTGATCATTCGCGACAGCAAGACGAGGCGACCATCTT 1005  
Db 1047 CCATGATCATCGCGCATCAAGAGGATGAGGCGACTCTCTT 1086

RESULT 7

US-09-801-852A-1  
; Sequence 1, Application US/09801852A  
; Patent No. US20020048781A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; APPLICANT: Chiang, Shu-Jen  
; APPLICANT: Jonathan, Basch  
; TITLE OF INVENTION: DIRECT PRODUCTION OF DESACETYLCEPHALOSPORIN C  
; FILE REFERENCE: ON0163  
; CURRENT APPLICATION NUMBER: US/09/801,852A  
; CURRENT FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/188,033  
; PRIOR FILING DATE: 2000-03-09  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 1716  
; TYPE: DNA  
; ORGANISM: Rhodosporidium toruloides  
US-09-801-852A-1

Query Match 5.8%; Score 88.2; DB 9; Length 1716;  
Best Local Similarity 55.2%; Pred. No. 2e-14;  
Matches 218; Conservative 0; Mismatches 168; Indels 9; Gaps 2;

QY 274 GAGGACTGCCTCACCATCAACGTGGTGGCGCGCGGACCAAGCGGGCGGCGCAACCTC 333  
DB 361 GAGGATTGCCTCTCTCAATGTCGTTGCCCGCGCGGCTGTCGAGGGCGACAATCTT 420

QY 334 CCGGTCAATGCTCTGGATCTTTTGGCGGTGGGTTTGAGATGGGAGCGCCACCACTCTTCCT 393  
DB 421 CCGGTCTCTGTACATTCACGAGGTGGCTACGCGCTTCGGCGATCGAGCACCGGCGC 480

QY 394 CCGCGCCAGATGTCACCAAGAGTGTCTCATGGGCAAGCCATCATCAAGTGGCGGTCA 453  
DB 481 GACTTTGCCGCTTCACCAAGCACAG-----GGAACCAAGATGGTCTGTGTAATCTCC 535

QY 454 ACTACCGTGTTCCTCGTGGGGTCTTCGTGGTGTGATGACATCAAGCGCGAGGCGCG 513  
DB 536 AGTACCGTCTCGGACGCTTTGGTTCTCTCGTGGCCAGCCATGAAGGATACGGTGTA 595

QY 514 GGAACCGCGGCTTGAAGGACAGCGTTTGGGCATGCAGTGGGTGGGAGACAACATTGCG 573  
DB 596 CGAACCGCGGCTTGTGTGACAGCAATTCGCCCTTCATGGGTTCACAGCACGTCGGA 655

QY 574 GPTTCGCGCGGCGACCGAGCAAGGTGACATCTTTGG-----CGAGCGGGCAGCATGTCG 629  
DB 656 AGTTCGCGCGAACCCTGATACGATTGGGGCGAGTCTGCAGCGCGAGGTCG 715

QY 630 TGTGTGCCACCTCATCTGGAACGCGGCGACAAC 664  
DB 716 TTATGAACCAGATCAATTGCGACGCGCGCAACACC 750

RESULT 8  
US-10-369-493-34501  
; Sequence 34501, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 34501  
; LENGTH: 1419  
; TYPE: DNA  
; ORGANISM: *Sphingomonas aromaticivorans*  
US-10-369-493-34501

Query Match 5.5%; Score 84.6; DB 16; Length 1419;  
Best Local Similarity 57.8%; Pred. No. 2e-13;  
Matches 196; Conservative 0; Mismatches 129; Indels 14; Gaps 2;

QY 274 GAGGACTGCCTCACCATCAACGTGGTGGCGCGCGGACCAAGCGGGCGGCGCAACCTC 333  
DB 220 GAGGATTGCCTCTATCTCAACGTATGAAGCCCGGTCCGCGCAAGCGGGAGGAACTG 279

QY 334 CCGGTCAATGCTCTGGATCTTTTGGCGGTGGGTTTGAGATCGGAGCGCCACCACTCTTCCT 393  
DB 280 CCGGTCAATGCTCTGGATCCACGCGGCGGTTTTCGTGAACGGAGG-----TTCCT 328

QY 394 CCGGCGCCAGATGTTCAACCAAGAGTGTGCTCATGGGCAAGCCATCATCAAGTGGCGGTCA 453  
DB 329 CCGGCGCATCTTATTCGGGCGAGAACTTCGCGCGCGAGCGGTCTGCTGTGAGCTGA 388

```
; SEQ ID NO 858
; LENGTH: 2312
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1763)..(1763)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-858

Query Match      4.0%; Score 61.2; DB 16; Length 2312;
Best Local Similarity 52.7%; Pred. No. 8.7e-07;
Matches 214; Conservative 0; Mismatches 173; Indels 19; Gaps 3;

QY 236 TGCAGTCCAAAGGTGTTCCAGGCGGTGCTTCCAGAGTGAGGACTGCTCACCATCAACG 295
Db 267 TGCAGGCCACCATCACCAGGAGACACCTACGGGGATGAAGACTGCTGTACCTCAACA 326
QY 296 TGGTGGCGCGCGCGGACCAAGCGGGCGCC---AACCTCCCGGTGATGCTCTGGATCT 352
Db 327 TTTGGGTGCCCCCAGGCGAGGAAGCAAGTCTCCCGGACCTCCCGGACCTGCTGATCT 386
QY 353 TTGGCGGTGGTGTGAGATCGGCAGC-----CCCAACATCTTCCCTCCCGCCCG 402
Db 387 ATGGAGGCGCTTCTCTCATGGGTCCGGCCATGGGGCAACTCTCTCAACAACATACCTGT 446
QY 403 ATGCTCACCAGAGTGTCTCATGGGCAAGCCATCATCCAGTGGCGGTCACTACCGTG 462
Db 447 ATGACGGCGAGGAGATGCCACACGCGGAACGTCATGCTGTGCTCACCTTCAACTACCGTG 506
QY 463 TTGGCTGTGGGGGTCTTGGCTGGTGTGATGATCAAGCGGAGGCGGAGCGGAACGCG 522
Db 507 TCGGCGCCCTTTGGGTTCCTCAGCACTGGGGAC-----GCCAATCTGCAGGTACTATG 560
QY 523 GCTTGAAGACACAGCGTTTGGGCATGCAGTGGGTGCGAGCAACAATTCGCGGTTCGCGG 582
Db 561 GCCTTCGGGATCAGCAGATGCCCATGCTTGGGTGAAGAGGAATATCGCGGCTTCGCGG 620
QY 583 GCGACCCGAGCAAGGTGACATCTTTGGGAGCGGCGGAGCATGTC 628
Db 621 GGGACCCCAACAACATCACGCTCTTCGGGGAGTCTGCTGGAGGTGC 666

RESULT 12
US-09-969-347-220
; Sequence 220, Application US/09969347
; Patent No. US20020115085A1
; GENERAL INFORMATION:
; APPLICANT: Eber, Reinhard
; TITLE OF INVENTION: Cancer
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-69
; CURRENT APPLICATION NUMBER: US/09/969,347
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/60/237,598
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,604
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 318
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 220
; LENGTH: 2428
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-969-347-220

Query Match      4.0%; Score 61.2; DB 9; Length 2428;
Best Local Similarity 52.7%; Pred. No. 8.8e-07;
Matches 214; Conservative 0; Mismatches 173; Indels 19; Gaps 3;

QY 236 TGCAGTCCAAAGGTGTTCCAGGCGGTGCTTCCAGAGTGAGGACTGCTCACCATCAACG 295
Db 344 TGCAGGCCACCATCACCAGGAGACACCTACGGGGATGAAGACTGCTGTACCTCAACA 403
```

```
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7341
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1476)
US-10-156-761-7341

Query Match      4.0%; Score 61.2; DB 15; Length 1476;
Best Local Similarity 53.0%; Pred. No. 7.7e-07;
Matches 215; Conservative 0; Mismatches 168; Indels 23; Gaps 3;

QY 216 GACGCACTCGACTTGGTGTGATCCAGTCCAAAGGTGTTCCAGGCGGTGTTCCCGCAGATGA 275
Db 144 GCGGCCCCAGGAGCGGGGATCCAGGCGCGCGCCCTGTGTGAGCGGCCACCGGCGA 203
QY 276 GGACTGCTCACCATCAAGTGTGCGGCGCGGCGGACCAAGCGGCGGCGCAACTCCC 335
Db 204 CGACTGGTGTGAGGTCAAC---GTCTGGACACCGGACCGGACCTGTGTGCGCGCGCC 260
QY 336 GGTATGCTCTGGATCTTTGGCGTGGTGTGAGATCGGACCGCCACCATCTTCCCTCC 395
Db 261 GGTGATGTTGTGATCTACGCGGTTGCTTACAGCTCGGCACTCGGCGAGCCCGGCTA 320
QY 396 CCGCCAGATGTTCAACAAGAGTGTCTCATGGGCAAGCCATCATCAGTGGCGGCTCAAC 455
Db 321 CGACGCTCGCGCATC-----GCCGCGACGCGGACGTCGTGTGTCTACCTCAAC 372
QY 456 TACCGTGTGCTGTGGGGTCTTGGCTGTGTGATGATCAATCAAGCGGCGGAGCGG 515
Db 373 TACCGTGTGCGCATCGAGGCTTCGCGCGGTCGAGCGGCTCCCGCC----- 420
QY 516 AACGCGGCTTCAAGACGAGCGTTTGGGCATGCGAGTGGGTGGCGAGCAACAATTGCGGG 575
Db 421 AACCGGCTGTCTCGACCAAGTTCGCGGCTTGGATGTTGTTGAGTACGAGAACATCAGCGG 480
QY 576 TTCCGCGCGGACCCGAGCAAGGTGACATCTTTGGGAGCGGCGGAG 621
Db 481 TTCCGCGGCGGACCCCGCGGCTACCGTCTTCGGGGAGTTCGCGCG 526

RESULT 11
US-10-264-049-858
; Sequence 858, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
```



```

; LOCATION: 2152..2184
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 2185..2217
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 2218..2250
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 2251..2283
; PUBLICATION INFORMATION:
; AUTHORS: Nilsson, Jeanette
; AUTHORS: Blackberg, Lars
; AUTHORS: Carlsson, Peter
; AUTHORS: Enerback, Sven
; AUTHORS: Hernell, Olle
; AUTHORS: Bjursell, Gunnar
; TITLE: cDNA cloning of human-milk
; TITLE: bile-salt-stimulated lipase and evidence for its
; TITLE: identity to pancreatic carboxylic ester hydrolase
; JOURNAL: Eur. J. Biochem.
; VOLUME: 192
; PAGES: 543-550
; DATE: Sept.-1990
; US-09-418-176-1

```

```

Query Match 4.0%; Score 61.2; DB 10; Length 2428;
Best Local Similarity 52.7%; Pred. No. 8.8e-07;
Matches 214; Conservative 0; Mismatches 173; Indels 19; Gaps 3;

Qy 236 TGCAGTCCAAAGTGTTCAGGCGGTCTTCCCGAGGTGAGGACTGCTCCACATCAACG 295
Db 344 TCGAGGCCACCATCACCAGGAGGACACCTAGCGGATGAAGACTGCTGTACCTCAACA 403
Qy 296 TGGTGGCGCGCGCGGACCAAGAGCGCGCGC-----AACCTCCGGTCAATGCTCTGGATCT 352
Db 404 TTTGGGTGCGCGCGGAGGAGCAAGTCTCCCGGACCTTCCCGGCTTATGATCTGGATCT 463
Qy 353 TTGGCGGTGGTGTGAGATCGGCAGC-----CCACCATCTTCCCTCCCGCCGAG 402
Db 464 ATGGAGGCGCCTTCTCTATGGGTCCGGCATGGGGCCAACTTCTCAACAACTACCTGT 523
Qy 403 ATGCTCACAAGAGTGTCTCATGCGGCAAGCCATCATCCAGTGGCGGTCAACTACCGTG 462
Db 524 ATGACGGCGAGAGATGCCACACGCGGAAGCTCATGCTGTGATCACTTCAACTACCGTG 583
Qy 463 TTGCTCTGTGGGTCTTGTGGTGTGATGATCAAGCCGAGGCGGAGCGGACGCGG 522
Db 584 TCGGCGCGCTTGGGTCTCTCAGCACTGGGAGC-----GCCAATCTGCCAGGTAACTATG 637
Qy 523 GCTTGAAGGACGAGGTGTGGCATGCGAGTGGTGGCGAGACAACTTCCCGGTTTCGGCG 582
Db 638 GCCTTCGGGATCAGACATGGCCATGCTTGGGTGAAGAGATATCGGGCTTCGGGG 697
Qy 583 GCGACCCGAGGAGTGATCATCTTTGGCGAGCGCGGAGCATGTCC 628
Db 698 GCGACCCCAACAACATCAGCTCTTCGGGGAGTCTGCTGGAGGTGC 743

```

```

RESULT 14
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262

```

```

; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 4.0%; Score 61.2; DB 15; Length 9025608;
Best Local Similarity 53.0%; Pred. No. 7.4e-06;
Matches 215; Conservative 0; Mismatches 168; Indels 23; Gaps 3;

Qy 216 GACGGCACTCGACTTGGTGTGATGCAGTCCAAAGTGTTCACAGCGGTGCTTCCCAAGTGA 275
Db 8791846 GCCGCCCCAGGAGACGGGATCCAGGGCGCGCGCCCTGCTGGACGCGCCACACGGCGA 8791787
Qy 276 GGACTGCTCACCATCAACGTGGTGGCGCGCGCGCGGACCAAGCGGGCGCGCAACTCTCC 335
Db 8791786 CGACTGGTGTACGGTCAAC--GTCTGGACACACCGGACCGCTGTGCGCGCGCGCC 8791730
Qy 336 GGTATGCTCTGGATCTTTGGCGGTGGTGTGAGATCGGACGCCACCATCTTCTCCCTCC 395
Db 8791729 GGTATGCTGTGATCTACGCGGTGCTTACAGCTGGCCACTCCGCGACGCCCGGTA 8791670
Qy 396 CGCCAGATGTACCAAGAGTGTCTCATGGCAAGCATATCCAGTGGCGGTCAAC 455
Db 8791669 CGAGCTCGCGCATC-----GCCGCGACGCGGACGTCGTCGTCACTCAAC 8791618
Qy 456 TACGCTGTGCTCGTGGGTCTTGTGGTGTGATGATCAAGCGGCGAGGCGGACGCGG 515
Db 8791617 TACGCTGTGCGCATCGAGGGCTTCGCGCGGTCTGACGCGCTCCGCGC----- 8791570
Qy 516 AACCGCGCTTGAAGGACCGCTTTCGGCATGCGAGTGGGTGGAGACAACTTCCCGGG 575
Db 8791569 AACCGCGCTCTGCTCGACCGCTCGGCGCTTGAATGGGTACGGGAGAACTACGCG 8791510
Qy 576 TTCGCGCGGACCGCGAGCAAGTGTACATCTTTGGCGAGCGCGGAG 621
Db 8791509 TTCGCGCGGACCGCGCGGCTCACCGTCTTCGGGGAGTCGGCG 8791464

```

```

RESULT 15
US-10-029-386-22811/c
; Sequence 22811, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEONICA-X-2
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22811
; LENGTH: 740
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF002993.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7

```



Search completed: August 5, 2004, 19:14:47  
Job time : 538 secs

Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2004, 15:42:35 ; Search time 2825 Seconds  
(without alignments)  
16194.287 Million cell updates/sec

Title: US-09-943-857-3  
Perfect score: 1532  
Sequence: 1 gccaccgcgaagctcgcca.....tgaccacccgttttctgtg 1532

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_pbg:\*

27: em\_gss\_vrl:\*

28: gb\_gsal:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	173.2	11.3	769	14	CF881106
C 2	173.2	11.3	824	14	CB907899
C 3	159.2	10.4	719	14	CF880958
C 4	159.2	10.4	774	14	CB907713
					CF881106 tric083xn
					CB907899 tric083xn
					CF880958 tric083xc
					CB907713 tric083xc

C	5	150.6	9.8	970	14	CF885800	CF885800
	6	117.4	7.7	1070	14	CF885724	tric083xc
	7	110	7.2	708	14	CF880550	tric082xn
	8	110	7.2	762	14	CB907475	tric082xn
	9	104.8	6.8	763	29	CG810164	CG810164
	10	88.6	5.8	681	14	CD054203	F5AAQ71TF
	11	86.4	5.6	449	14	CD054203	H002B22S
	12	85.8	5.6	620	10	ADW056446	ADW056446
	13	85	5.5	559	12	AW790714	D00142-R
	14	81.4	5.3	1800	9	BM361770	BM361770
	15	75.2	4.9	781	29	AA415091	AA415091
	16	67	4.4	875	12	CG811743	CG811743
	17	63.6	4.2	931	12	BI952106	BI952106
	18	62.2	4.1	751	14	BI488594	BI488594
	19	61.4	4.0	604	14	CF342195	CF342195
	20	61.2	4.0	780	12	CD056060	CD056060
	21	61.2	4.0	902	14	BI753192	BI753192
	22	60	3.9	611	13	CK249907	CK249907
	23	60	3.9	941	13	BQ225068	BQ225068
	24	60	3.9	1138	12	BM545507	BM545507
	25	60	3.9	1194	14	CD014076	CD014076
	26	60	3.9	1305	14	CD014075	CD014075
	27	60	3.9	1364	14	CD014074	CD014074
	28	60	3.9	1472	14	CD014072	CD014072
	29	60	3.9	1496	14	CD014071	CD014071
	30	60	3.9	1585	14	CD014073	CD014073
	31	60	3.9	1723	29	AY407736	AY407736
	32	60	3.9	1723	29	AY407737	AY407737
	33	60	3.9	1745	11	BC036813	BC036813
	34	60	3.9	2080	11	BC001541	BC001541
	35	60	3.9	2187	11	BC026315	BC026315
	36	59.6	3.9	657	14	CF523157	CF523157
	37	59.6	3.9	671	14	CF524855	CF524855
	38	59.6	3.9	681	14	CF524540	CF524540
	39	59.6	3.9	698	14	CF523487	CF523487
	40	59.6	3.9	698	14	CF343579	CF343579
	41	59.6	3.9	722	14	CF525122	CF525122
	42	59.6	3.9	723	14	CF524239	CF524239
	43	59.6	3.9	724	14	CF525075	CF525075
	44	59.6	3.9	755	14	CF523991	CF523991
	45	59.6	3.9	757	14	CF374578	CF374578

#### ALIGNMENTS

#### RESULT 1

CF881106/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CF881106 769 bp mRNA linear EST 31-OCT-2003  
tric083xn18.b1 T.reesei mycelial culture, Version 6 October 2003  
Hypocrea jecorina cDNA clone tric083xn18, mRNA sequence.

CF881106

CF881106.1

GI:38135788

Hypocrea jecorina (anamorph: Trichoderma reesei)

Hypocrea jecorina

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Hypocremycetidae; Hypocreales; Hypocreaceae; Hypocrea.

1 (bases 1 to 769)

Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D.,

Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and

Dean, R.A.

Analysis of the protein processing and secretion pathways in a

Trichoderma reesei EST dataset

Unpublished (2003)

Contact: Ralph A. Dean

Fungal Genomics Laboratory

North Carolina State University

Campus Box 7251, Raleigh, NC 27695, USA

Tel: 919-513-0020

Fax: 919-513-0024

Email: ralph.dean@ncsu.edu

Seq primer: Lt-F1 primer.

```
FEATURES          Location/Qualifiers
  source
    1..769
      /organism="Hypocrea jecorina"
      /mol_type="mRNA"
      /strain="QM6a"
      /db_xref="taxon:51453"
      /clone="trico83xn18"
      /dev_stage="mycelia"
      /clone_lib="T.reesei mycelial culture, Version 6 October
      2003"
      /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
      culture grown from 24 hrs to 6 days with varying Carbon
      and Nitrogen sources and concentrations."

ORIGIN
  Query Match      11.3%; Score 173.2; DB 14; Length 769;
  Best Local Similarity 60.3%; Pred. No. 2.1e-20;
  Matches 364; Conservative 0; Mismatches 223; Indels 17; Gaps 4;

QY 257 CGGTGCTTCCCGAGAGTGGAGTGCCTCACCATCAACAGTGGTGGCGCGCGGACCA 316
DB 663 CGGAGGACCAACAGTGGGAAAGTCCCTTGACATCAATCACTCTGTCGCGAGCGCTTTCCTCATTTGGCG 604
QY 317 AGCGGGGCGCAACCTCCCGGTTCATGCTCTGGATCTTTGGCGGTGGGTTCGATTCGGCA 376
DB 603 AACCGGGGAGCGCTGCGGGTGTCTTTTACATCTTCGAGCGGCTTTCCTCATTTGGCG 544
QY 377 GCGCCA-----CCATCTTCCCTCCCGCGCAGATGTCACCAAGATGTCATCGGCA 430
DB 543 GGACGTGGCTGCCCTGAACAGACCCCTCCAGCTGTCGACAGCGGGCGTCCGCTTCGGCA 484
QY 431 AG-CCATCATCACTGCGCGTCACTACGCTGTCCTGCGGGGTTCCTGGCTGGT 489
DB 483 AGCCCTTCATCTTCGCGGCGTCACTACCGGTGCGCGGTGGGGTTCATGCTGGG 424
QY 490 ATGACATCAAGCGCGAGCGGAGCGCGGCTTGAAGGACCAAGCGTTTGGCGATGC 549
DB 423 AGGAGATTCTCAATGAGGAGCGCAACGCGGCTTGTTCACGAGCGATGGGCTCG 364
QY 550 AGTGGGTGGGACAGCAATTCGCGGTTCGCGGGGAGCGCGGAGAGTGCATTTGG 609
DB 363 AATGGGTGGCGGATCAATGAGGCTTTGGCGGGAACCTCGTCCGAGGTGACCATCTGG 304
QY 610 CG-----AGCGGGGAGCATGTCGCTGTTGTCACCTCATCTGGAACAGCGGCAACA 665
DB 303 GGCAGTCCCTGGCTCCATTCGCTGGGACAGCTCGTGTGTATGATGGGATGCCA 244
QY 666 CGTACAAGGCGCAAGCGGTGTTTCGCGCGGGGCATCATGCAAGGAGCCATFGTGGCG 721
DB 243 CATACAACGACAAGCTCTGTTTCGCGCGCGCCATCATGAATCTCGGCTCCGTGACGCGCG 184
QY 722 --GACCGGTGACGCGACGTACGCAACGAGATCTACGACCTTCTTGTCTCGAGTGTG 779
DB 183 TCGATCCGCTGACTCGGAAAGGCGAAGCGGCTGATCAGCATGTCGTCGAGAGGCTG 124
QY 780 GCTGTGGGAGCGCGACGCAAGCTCGCTGCTTGGCAGTGGAGCGGACACCTTGCTCG 839
DB 123 GATGGACGGGAGCGAGCAACCTCGGTGTCTCGGAATCTGACCAAGGAGCTTCG 64
QY 840 ATGC 843
DB 63 CCGC 60

RESULT 2
CB907899/c      824 bp      mRNA      linear      EST 02-JUL-2003
LOCUS          trico83xn18 T.reesei mycelial culture, Version 3 april Hypocrea
DEFINITION     jecorina cDNA clone trico83xn18, mRNA sequence.
ACCESSION     CB907899
VERSION       CB907899.1 GI:30122557
KEYWORDS      EST.
SOURCE         Hypocrea jecorina (anamorph: Trichoderma reesei)

ORGANISM       Hypocrea jecorina
REFERENCE      1 (bases 1 to 824)
AUTHORS       Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
              Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,
              Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
              Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
              Transcriptional regulation of biomass-degrading enzymes in the
              filamentous fungus Trichoderma reesei
              J Biol. Chem. 278 (34), 31988-31997 (2003)
              22803314
              MEDLINE
              12788920
              PUBMED
              COMMENT
                Contact: Pamela K. Foreman
                Genencor Intl.
                925 Page Mill Road, Palo Alto, CA 94304, USA
                Tel: (650) 846-7635
                Fax: (650) 621-7817
                Email: Pforeman@genencor.com
                Seq primer: LT-F1 primer.
                Location/Qualifiers
                  1..824
                    /organism="Hypocrea jecorina"
                    /mol_type="mRNA"
                    /strain="QM6a"
                    /db_xref="taxon:51453"
                    /clone="trico83xn18"
                    /dev_stage="mycelia"
                    /clone_lib="T.reesei mycelial culture, Version 3 april"
                    /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
                    culture grown from 24 hrs to 6 days with varying Carbon
                    and Nitrogen sources and concentrations."

ORIGIN
  Query Match      11.3%; Score 173.2; DB 14; Length 824;
  Best Local Similarity 60.3%; Pred. No. 2.1e-20;
  Matches 364; Conservative 0; Mismatches 223; Indels 17; Gaps 4;

QY 257 CGGTGCTTCCCGAGAGTGGAGTGCCTCACCATCAACAGTGGTGGCGCGCGGACCA 316
DB 718 CGGAGGACCAACAGTGGGAAAGTGCCTTGACATCAATCACTCTGTCGCGAGCGCACGC 659
QY 317 AGCGGGGCGCAACCTCCCGGTTCATGCTCTGGATCTTTGGCGGTGGGTTCGATTCGGCA 376
DB 658 AACCGGGGAGCGCTGCGGGTGTCTTTTACATCTTCGAGCGGCTTTCCTCATTTGGCG 599
QY 377 GCGCCA-----CCATCTTCCCTCCCGCGCATGTCACCAAGAGTGTCTCATGGCA 430
DB 598 GGCAGTCCGCTGCCCTTGACGACCCCTCCAGTCTGTCAGAGCGGCTGCTTCGGCA 539
QY 431 AG-CCATCATCACTGCGCGGTCAACTACCGGTTCCTTCGCTGGGGGTTCCTGGCTGGT 489
DB 538 AGCCCTTCATCTTCGCGCGGTCAACTACCGGTTCGCGGGGTTCATGCTGGGG 479
QY 490 ATGACATCAAGCGCGAGCGGAGCGCGGCTTGAAGGACCAAGCGTTTGGGCTGC 549
DB 478 AGGAGATTCTCAATGAGGAGCGCAACGCGGCTTGTTCACCAAGCGCATGGGCTCG 419
QY 550 AGTGGGTGGGACAGCAATTCGCGGTTCGCGGGGAGCGCGGAGCGAGTGCATCTTTGG 609
DB 418 AATGGGTGGCGCATCAATGAGGCTTTCGCGGAGCTTCGTCGAGGTGACCATCTGG 359
QY 610 CG-----AGCGGGGAGCATGTCGCTGTTGTCACCTCATCTTGGAAACGAGCGGCAACA 665
DB 358 GGCAGTCCGCTGGCTCCATTTCCGTCTGGGAGCAGCTCGTGTGTATGATGGGATGCCA 299
QY 666 CGTACAAGGCGCAAGCGGTGTTTCGCGCGCGGATCATGCAAGGAGCATGTGTGGCG 721
DB 298 CATACAACGACAAGCTCTGTTTCGCGCGCGCCATCAATGAATCTCGGCTCCGTGACGCGCG 239
QY 722 --GACCGGTGAGCGGACGTACGCAACGAGATCTTACGACCTTCTTGTCTCGAGTGTG 779
DB 238 TCGATCCGCTCGACTCGGAAAGGCGAAGCGGCTGTATCAGCATGTCGTCGAGAGGCTG 179
```

QY 780 GCTGTGGCAGCGCCAGCGCAAGCTCGCGTCTTGGCCAGTGGCGAGCGACACCTTGTCTCG 839  
 |||||  
 Db 178 GATGCGACGCGCAGCGACGACACCTGGGTGTCTCCGGAACTTGACCAACGACGATTCG 119  
 |||||  
 QY 840 ATGC 843  
 |||||  
 Db 118 CCGC 115

RESULT 3  
 CF880958 719 bp mRNA linear EST 31-OCT-2003  
 LOCUS  
 DEFINITION  
 trico83xc07.b1 T.reesei mycelial culture, Version 6 October 2003  
 Hypocrea jecorina cDNA clone trico83xc07, mRNA sequence.  
 CF880958  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Hypocrea jecorina (anamorph: Trichoderma reesei)  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
 1 (bases 1 to 719)  
 Diener S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D.,  
 Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and  
 Dean, R.A.  
 Analysis of the protein processing and secretion pathways in a  
 Trichoderma reesei EST dataset  
 Unpublished (2003)  
 Contact: Ralph A. Dean  
 Fungal Genomics Laboratory  
 North Carolina State University  
 Campus Box 7251, Raleigh, NC 27695, USA  
 Tel: 919-513-0020  
 Fax: 919-513-0024  
 Email: ralph.dean@ncsu.edu  
 Seq primer: IT-F1 primer:  
 Location/Qualifiers  
 1..719  
 /organism="Hypocrea jecorina"  
 /mol\_type="mRNA"  
 /strain="QM6a"  
 /db\_xref="taxon:51453"  
 /clone="trico83xc07"  
 /dev\_stage="mycelia"  
 /clone\_lib="T.reesei mycelial culture, Version 6 October  
 2003"  
 /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial  
 culture grown from 24 hrs to 6 days with varying Carbon  
 and Nitrogen sources and concentrations."

ORIGIN  
 Query Match 10.4%; Score 159.2; DB 14; Length 719;  
 Best Local Similarity 61.4%; Pred. No. 5e-18;  
 Matches 349; Conservative 0; Mismatches 198; Indels 21; Gaps 5;

QY 459 CCGTGTGCTCGTGGGGGTCTTGGCTGGTGATGATCAAGCCCGAGCGCGGCGGAAC 518  
 |||||  
 Db 1 CGCGTCGCGCGGTCTTGTCTTCTGGCGCGCGGAGATCTCGGCAACGGCAGCGCGAAC 60  
 |||||  
 QY 519 GCCGCTTGAAGCAGCAGGCTTTGGCATGTCAGTGGTGGCAGACAACATNTCCCGGTTTC 578  
 |||||  
 Db 61 CTCGGGCTGCTGGACCGCGATCGGCTTGCAGTGGTGGCGGACAAATCGCGGCTTT 120  
 |||||  
 QY 579 GCGCGCGCCCGCAGCAAGTGA-CATCTTTGGCGAG--GCCGGCAGCATGTCGGTGTG 634  
 |||||  
 Db 121 GCGCGGACCCGGAAGAAGTGACCATCTGCGGCGAGTCGCGGCGGCCCATGTCGCTTAC 180  
 |||||  
 QY 635 TGCCACCTCATCTGGAACGACCGCGCAACAGCTACAAAGGCGCAAGCCGTTGTTCGCGGG 694  
 |||||  
 Db 181 GACCAGATGGTGTGTACGCGCGCAACGCCAGTACACGCGCAAGCCCTGTTCGCGCG 240  
 |||||  
 QY 695 GGCATCATGCA-----GGGAGCCATGTTGTCGGGACCCCGGTGGACGGCACTGACGGCAAC 748  
 |||||

Db 241 GCCATCATGAACCTTGCAACGGCGTCCCGCGAGCGCTCGACTCGCCCAAGCGCGAG 300  
 |||||  
 QY 749 GAGATCTACACACTCTTTGTCTGAGTGTGGTGTGGTGTGGAGCGCCAGCGCAAGCTCGCG 808  
 |||||  
 Db 301 GCCATCTACGACAAACGTCGTGAAAAAGGCGCGGTGACAGCGCTCCGCGCACTCGCTCGCG 360  
 |||||  
 QY 809 TGCTTGGCAGTGGAGCGACAC--TTGCTCGATGCCACCAACACACTCTCTGGGTTTC 865  
 |||||  
 Db 361 TGCTCGCGCAGTGGCTTACGCAAGTTCTTCCGCCCTCAACGTGGCGCGCGCATC 420  
 |||||  
 QY 866 TTGGGTACTCTCTGTTGGGTTGTACTTC-----CGGCCCGAGCGCAAGAACATCA 917  
 |||||  
 Db 421 CTGTCTTACAGCTCCCTGGGGGCTGTCTTACCTTACCTGCGCGCGCGAGCGGTCTCTCTCC 480  
 |||||  
 QY 918 CCGATGACATGTACAAGTTGTCGCGGACGCGGAAGTATGCAAGCGTTCCCGTGATCATTTG 977  
 |||||  
 Db 481 CGGACAGCCCCGAAAGCGCTCTCGACAGCGGAGGTACCAAGCGCTGCCCCATGATTGCCG 540  
 |||||

QY 978 GCGACCAAGACGAGCGGCGACCATCTT 1005  
 |||||  
 Db 541 GCGACGAGGAGGAGGCGCGCTCTT 568  
 |||||

RESULT 4  
 CB907713 774 bp mRNA linear EST 02-JUL-2003  
 LOCUS  
 DEFINITION  
 trico83xc07 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico83xc07, mRNA sequence.  
 CB907713  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Hypocrea jecorina (anamorph: Trichoderma reesei)  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
 1 (bases 1 to 774)  
 Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,  
 Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,  
 Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,  
 Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.  
 Transcriptional regulation of biomass-degrading enzymes in the  
 filamentous fungus Trichoderma reesei  
 J. Biol. Chem. 278 (34), 31988-31997 (2003)  
 22803314  
 12788920  
 Contact: Pamela K. Foreman  
 Genencor Intl.  
 925 Page Mill Road, Palo Alto, CA 94304, USA  
 Tel: (650) 846-7635  
 Fax: (650) 621-7817  
 Email: Pforeman@genencor.com  
 Seq primer: IT-F1 primer:  
 Location/Qualifiers  
 1..774  
 /organism="Hypocrea jecorina"  
 /mol\_type="mRNA"  
 /strain="QM6a"  
 /db\_xref="taxon:51453"  
 /clone="trico83xc07"  
 /dev\_stage="mycelia"  
 /clone\_lib="T.reesei mycelial culture, Version 3 april"  
 /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial  
 culture grown from 24 hrs to 6 days with varying Carbon  
 and Nitrogen sources and concentrations."

ORIGIN  
 Query Match 10.4%; Score 159.2; DB 14; Length 774;  
 Best Local Similarity 61.4%; Pred. No. 6e-18;  
 Matches 349; Conservative 0; Mismatches 198; Indels 21; Gaps 5;

QY 459 CGTGTGCTCGTGGGGGTCTTGGCTGGTGATGATCAAGCGCGGCGGCGGAAC 518  
 |||||



Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph\_dean@ncsu.edu  
Seq primer: LT-F1 primer.  
Location/Qualifiers  
1. .1070  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric083xc05"  
/dev\_stage="mycelial"  
/clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

FEATURES source

```

Seq primer: LT-F1 primer
Location/Qualifiers
1..708
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="tric082xh02"
/dev_stage="mycelia"
/clone_lib="T.ressei mycelial
2003"

```

```

/clone_lib="T.reesei mycelial culture, Version 6 October
2003"
/notes="Vector: PREP3Y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations "
```

```

/dev stage="mycelia"
/clone_lib="T.Reesei mycelial culture, Version 6 October
2003"
/note="Vector: pREP3Y; Site_1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

```

Query Match 7.2%; Score 110; DB 14; Length 708;  
Best Local Similarity 60.3%; Pred. No. 2.7e-09;  
Matches 216; Conservative 0; Mismatches 140; Indels

ORIGIN

Query Match	7.7%;	Score 117.4;	DB 14;	Length 1070;
Best Local Similarity	70.0%;	Pred. No. 1.6e-10;		
Matches 187;	Conservative	0;	Mismatches 76;	Indels 4;
			Gaps	2;

QY	459	CGTGTTCCTCTGCTGGGGGTTCTTGCTGTGTGATGACATCAAGCCGAGCGCAGCGGAAC	518
Db	1	CGCGTCGCCCGCTTCGGCTTCCTGGGGGGGCCGAGATCTGCGCAACGGCAGCGCGAAC	60
QY	519	GCCGGCTTGAAGACCACAGCGCTTTGGGCATCGTGGGTGGCAGACACATNTGCCGGTTC	578
Db	61	CTCGGGCTGCTGAGCCAGCGCATTGGGCTTCAGTGGGTGGCGAACATCGCGGCGCTTT	120
QY	579	GGCGGCGACCCGAGCAAGGTGA-CATCTTTGGCGAG---GGCGGACGATGTCGCTGTG	634
Db	121	GGCGGCGACCCGACCAAGGTGACCATCTGGGCGAGTCGCGGGCGCCATGTCGCTCTAC	180
QY	635	TGCCACTCATCTGGAAACGACGGCGCAACACCTGTACAGGGCGACCGTGTGTTCCGGCG	694
Db	181	GACCAATGGTGTGTACGGCGGCGAACGCCACCGTACACCGCAAGCCCTGTTCCCGGC	240
QY	695	GGCATCATGAGGAGCCATGTGCCG	721
Db	241	GCCATCATGAATCTCTGGCAGCGCGCG	267

## RESULT 7

CF880590 708 bp mRNA linear EST 31-OCT-2003  
tric082h202.bl T.reesei mycelial culture, 6 October 2003  
Hypocrea jecorina cDNA clone tric082h202, mRNA sequence.

ACCESSION  
CF880590  
CF880590.1 GI:38135272  
EST.

SOURCE	ORGANISM
Hypocrea jecorina	(anamorph: Trichoderma reesei)
Hypocrea jecorina	

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocorymbales; Hypocreales; Hypocreaceae; Hypocrea.  
1 (bases 1 to 708)

Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D., Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and Dean, R.A.

REFERENCE

AUTHORS

TITLE	JOURNAL
Analysis of the protein processing and secretion pathways in a <i>Trichoderma reesei</i> EST dataset	Unpublished (2003)

COMMENT:  
Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: [ralph.dean@ncsu.edu](mailto:ralph.dean@ncsu.edu)

RESULT 8  
CB907475

LOCUS	CB907475	762 bp	mRNA	linear	EST 02-JUL-2003
DEFINITION	trico82xh02 T.reesei mycelial culture, Version 3 april Hypocrea jecorina cDNA clone trico82xh02, mRNA sequence.				
ACCESSION	CB907475				
VERSION	CB907475.1	GI:30122133			
KEYWORDS	FST				

SOURCE	ORGANISM
Hypocrea jecorina (anamorph: Trichoderma reesei)	
Hypocrea jecorina	
Hypocrea jecorina	

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

1 (bases 1 to 762)

Foreman, P. K., Brown, D. E., Dankmeyer, L., Dean, R., Diener, S.,  
Dunn-Coleman, N. S., Goedegebuur, F., Houfek, T. D., England, G. J.,  
Kelley, A. S., Meerman, H. J., Mitchell, T., Mitchinson, C.,  
Olivares, H. A., Teunissen, P. J., Yao, J. and Ward, M.

REFERENCES

AUTHORS

**TITLE** Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*

**JOURNAL** J. Biol. Chem. 278 (34), 31988-31997 (2003)

**MEDLINE** 22803314  
**PUBMED** 12788920

COMMENT  
Contact: Pamela K. Foreman  
Genencor Int'l

925 Page Mill Road, Palo Alto, CA 94304, USA  
Tel: (650) 846-7635  
Fax: (650) 621-7817  
Email: [pfloren@genencor.com](mailto:pfloren@genencor.com)

Seq primer: LT-F1 primer.

Location/Qualifiers

source

1. .762

/organism="Hypocrea jecorina"

/mol\_type="mRNA"

/strain="QM6a"

/db\_xref="taxon:51453"

/clone\_lib="T.reesei"

/dev\_stage="mycelia"

/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial

culture grown from 24 hrs to 6 days with varying Carbon

and Nitrogen sources and concentrations."

http://Fusariumvirguliform.siu.edu

Seq primer: TGTAAACGACGGCCAGT

Class: BAC ends.

Location/Qualifiers

source

1. .763

/organism="Fusarium virguliforme"

/mol\_type="genomic DNA"

/cultivar="Monticello"

/db\_xref="taxon:232082"

/clone="KMFv5K21"

/clone\_lib="LargeInsertGenomicLibrary"

/note="Organ: Hyphae; Vector: pINDIGOBAC5; A single spore

derived culture was used. Hyphae were grown in an

incubator for four days. Nuclei were isolated and embedded

in agarose, restriction digested with Hind III. Large size

DNA fragments were ligated in vector pINDIGOBAC5 and

electro-transformed into DH10B cells."

FEATURES

source

1. .762

/organism="Hypocrea jecorina"

/mol\_type="mRNA"

/strain="QM6a"

/db\_xref="taxon:51453"

/clone\_lib="T.reesei"

/dev\_stage="mycelia"

/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial

culture grown from 24 hrs to 6 days with varying Carbon

and Nitrogen sources and concentrations."

FEATURES

source

1. .763

/organism="Fusarium virguliforme"

/mol\_type="genomic DNA"

/cultivar="Monticello"

/db\_xref="taxon:232082"

/clone="KMFv5K21"

/clone\_lib="LargeInsertGenomicLibrary"

/note="Organ: Hyphae; Vector: pINDIGOBAC5; A single spore

derived culture was used. Hyphae were grown in an

incubator for four days. Nuclei were isolated and embedded

in agarose, restriction digested with Hind III. Large size

DNA fragments were ligated in vector pINDIGOBAC5 and

electro-transformed into DH10B cells."

FEATURES

source

1. .762

/organism="Hypocrea jecorina"

/mol\_type="mRNA"

/strain="QM6a"

/db\_xref="taxon:51453"

/clone\_lib="T.reesei"

/dev\_stage="mycelia"

/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial

culture grown from 24 hrs to 6 days with varying Carbon

and Nitrogen sources and concentrations."



**ORGANISM**

*Hordeum vulgare*  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
1 (bases 1 to 681)  
Zierold,U. and Schweizer,P.  
Barley ESTs from pathogen-attacked leaf epidermis  
Unpublished (2003)  
Contact: Patrick Schweizer  
Transcriptome Analysis, Cytogenetics Department  
Institute of Plant Genetics and Crop Plant Research (IPK)  
Corrensstr. 3 D-06466 Gatersleben, Germany  
Tel: 0049 (0)39482-5660  
Fax: 0049 (0)39482-5595  
Email: schweiz@ipk-gatersleben.de  
Insert Length: 681 Std Error: 0.00  
Plate: 2 row: B column: 22  
Seq primer: SK.

**FEATURES**

source	Location/Qualifiers
	1..681
	/organism="Hordeum vulgare"
	/mol_type="mRNA"
	/cultivar="Ingrid BC mlo-5"
	/db_xref="GABI:704066"
	/db_xref="taxon:4513"
	/clone="HO02B22"
	/tissue_type="leaf epidermis, 6 h and 24 h post inoculation with Blumeria graminis"
	/dev_stage="7 d after germination"
	/lab_host="XL10-Gold"
	/clone_lib="HO"
	/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); Approximately 5 % of the clones correspond to cDNA from the fungi B. graminis hordei and tritici, respectively. Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 1.2 kb"

**ORIGIN**

Query Match	5.8%; Score 88.6; DB 14; Length 681;
Best Local Similarity	54.8%; Pred. No. 1.6e-05;
Matches 223; Conservative	0; Mismatches 174; Indels 10; Gaps 2
<hr/>	
QY 420	GCTCATGGCGAAGCCATCATCAAGCCGAGGGACGGGACGCCGTTCGAAGCACAGCGT 479
Db	67 GATAATGGAAACCATTGTCTACATGCTGTCAAATCTCGTTGGTGCTTTGGATGG 126
<hr/>	
QY 480	TTVGCTGGTGATGATCAATCAAGCCGAGGGACGGGACGCCGTTCGAAGCACAGCGT 539
Db	127 CTCGCTGGAAGAGGAGTATTGCATGCTGGGCTTACAAATTTGGGACACTACGATCAGATT 186
<hr/>	
QY	540 TTGGGCATGCAGTGGGTGGCAGACAACATTCGCGGGTTCGGCGGCGACCCGAGCAAGTG 599
Db	187 CTTGCTCTGCATGGTTTCAAGACAACATCTGAATTCGGTGAGACCCAGAACAGTT 246
<hr/>	
QY 600	ACATCTTTGGCGA----GGCGGGCAGCATGTCCGTGTTGTGCCACCTCATCTGAACGAC 655
Db	247 ACTCTCTTTGGAACATCTGCTGGTGGATTCTACATGGAACCTATTGACTGCTAAGAT 306
<hr/>	
QY	656 GGCGACAACAGTACAGGGCAAGCGGTGTTTCGCGCGGCATCATCGAGGAGCCATG 715
Db	307 GGTAACCTACGATCAACAGCAAGACGCTCTGTTCGAGCTGCCATTATGGACAGTGGCAGT 366
<hr/>	
QY 716	GT-----GCCGGACCCGGTGGACGGCAGTACGGCAACGAGATCTACGACCTCTTTGTC 769
Db	367 GTCATCCCAACAGACCACGCTGACTGGCCCTAAAGCCAGAACATTTTTAACACTGTCGT 426
<hr/>	
QY 770	TCCAGTGTGCTGTGGCAGCGCCGAGCAAGAAGCTCGCGTCTTTGGC 816

Qy	450	GTCAACTACCGTGTGTGGCTCGTGGGGTCTTTGGCTGGTGAAGACATCAAGCCGAGGCG	509
Db	250	GTCAACTATCGTCTTGGTGTCTTTGGATGGCTCGCTGGAAAGGAGGTATTGGATGCTGGG	309
Qy	510	AGCGGGAAACGCGCGCTTTGAAGGACACAGCGTTTGGGCATGCAAGTGGGTGGCAGACAAACATT	569
Db	310	CTTACAAATTTGGGACACTACGATCAGATTCTTGCTCTGCAATGGGTTCAAGACAACATT	369
Qy	570	GCCGGGTTTCGCGCGGACCCGAGCAAGGTGACATCTTTGG	609
b	370	GCTAAGTTTCGGTGGAGACCCAGCAAGGTTACTCTCTTTG	409

RESULT 12	AW790714	620 bp	linear	EST 01-MAY-2001
LOCUS	AW790714			
DEFINITION	D00142-R Lambda Zap, <i>Stratagene Blumeria graminis</i> f. sp. hordei cDNA clone D00142 similar to linase 2 precursor, mRNA sequence.			

```

FEATURES
source
FOUR-NO.
Location/Qualifiers
1..620
/organism="Blumeria graminis f. sp. hordei"
/mol_type="rRNA"
/db_xref="taxon:62688"
/clone="D00142"
/cell_type="conidia"
/lab_host="Hordeum vulgare"
/clone_lib="Lambda Zap, Stratagene"

```

Query Match	5.6%;	Score 85.8;	DB 10;	Length 620;
Best Local Similarity	51.8%;	Pred. No. 4.7e-05;		
Matches 302;	Conservative	0;	Mismatches 262;	
			Indels 19;	Gaps 4

Db	275	ACCTTCGCTACACGACAAGCCCTCTGTTCCGAGCTGCATCATGACGACAGTGGAAAGTGTTA	334
Qy	719	-----CCGGACCCCGGTGGACGGCACAGTACGGCAACAGAGATCTACGACCTCTTTGTCTCGA	773
Db	335	TCCCAACTGACCTCCCTCGCGATTGTCCCAAGGCCCAAAACATCTTCAACACTGTCTGTCGTG	394
Qy	774	GTGCTGCTCTGGCAGCGCCAGCGACGACAAGCTCGCGTCTTGGCGAGT--GCAGACGACACA	830
Db	395	CTGAGAGCTGTGAAGAGCTGCTGACAAAGATTGACTGCCTCCGAGCTCTGCCATACGACC	454
Qy	831	CCTTGTCTGATGCCAACCAACAACACTCCTCGGTTCTTGGCGTACTCCTCGTTGGGGTTGT	890
Db	455	AATTCTCCAACTAGTACATCTGTGCGCTGCTCTCTTCGACTACCCCGTCTGTGCGTCTTT	514
Qy	891	ACTCC-----CGGCCCGACGGAAGAACATACCGATGACATGTACAAAGTTGTTGGCGG	944
Db	515	CATACATTCCACGATACGCGGTGATTGTTGTGACAGATTCTGCCGACAACTTTGTGAGG	574
Qy	945	ACGGCAAGTATGCAAGCGTTTCCCGTGATCATTTGGCGACCGAA	987
Db	575	CTGGAAGTATGCCAAGTTCCTTACATTGCGGTGACCAAGAA	617

RESULT 13  
BM361770  
LOCUS BM361770 559 bp mRNA linear EST 01-JAN-2003  
DEFINITION A01032-R Appressorium stage EST library of *Blumeria graminis* f. sp. hordei *Blumeria graminis* f. sp. hordei cDNA similar to Lipase 2 precursor. mRNA sequence.

FEATURES	source
<p> POLIA=NC.  Location/Qualifiers  1..559  /organism="Blumeria graminis f. sp. hordei"  /mol_type="mRNA"  /db_xref="taxon:62688"  /tissue_type="Appressorium forming conidia"  /clone_lib="Appressorium stage EST library of Blumeria graminis f. sp. hordei"  /note="Vector: pBluescript II XR, Stratagene" </p>	

Query Match	5.5%	Score 85;	DB 12;	Length 559;
Best Local Similarity	53.5%;	Pred. No. 6.3e-05;		
Matches 251;	Conservative	0;	Mismatches 205;	Indels 13;
Gaps				

QY 544 GCATGCACTGGTGGCAGACAAATTTGCGGGTTCGGGGGACCCGAGCAAGGTGACAT 603  
 Db 174 CTCACAAATGGTTCACGATACATCGCAATTCGGTGGTATCCAGAAAGGTACTC 233  
 QY 604 CTTTGGGCGA-----GCGGGGACATGTCGGTGTGTCACCTCATCTGGAACAGCGCG 659  
 Db 234 TTTTCGGTACTTTCGAGGTGTTATCTCCACATGGAACCTGTTGACTGCCAAGACGGTA 293  
 QY 660 ACAAACAGTCAAGGGCAAGCGTGTTCGCGGGGATCATCGAGGAGCCATGTT-- 717  
 Db 294 ACCTTCGCTACAACGACAAGCGCTGTTCGGAGCTGCCATCATGACAGTGGAGTGT 353  
 QY 718 ----GCGGGACCGGTGACGACGACGTCAGCAACAGAGATCTACGACCTTTTCTCGA 773  
 Db 354 TCCCACTGACCTTCGCGATTTGTCGAGGCCCAAAACATCTTCAACACTGTCTTCGTG 413  
 QY 774 GTGTGCTGTGGCAGCGCCAGCAAGCTCGCGTCTTTCGCGAGT---GCGAGCGACA 830  
 Db 414 CTGGAGGCTGTGAAGGAGCTGCTGACAAATGTAATGCTCCCGAGCTTCGCCATACGACC 473  
 QY 831 CTTTGTCTGATGCAACCAACACTCTCTGGGTTCTTGGGCTACTCTCTC 879  
 Db 474 AATTCTCCCAAGTAGTACATCTGTGCTGTGCTCTCTTCGACTACCCGTC 522

## RESULT 14

AA415091 1800 bp mRNA linear EST 09-DEC-1999  
 LOCUS Mg0026 RCW Lambda Zap Express Library Magnaporthe grisea cDNA clone  
 DEFINITION RCW26 similar to Triacylglycerol Lipase (E.C.3.1.1.3), mRNA

## ACCESSION

AA415091 GI:2537256

## VERSION

EST.

## KEYWORDS

SOURCE

ORGANISM

Magnaporthe grisea (anamorph: Pyricularia grisea)

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.

1 (bases 1 to 1800)

Wu, S.-C., Bernstein, B.D., Darvill, A.G. and Albersheim, P.

Expressed sequence tags of the rice blast fungus grown on rice cell

walls

Unpublished (1997)

Contact: Sheng-Cheng Wu

CCRC

University of Georgia

220 Riverbend Road, Athens, GA 30602-4712, USA

Tel: 706 542 4446

Fax: 706 542 4412

Email: wusc@bscr.uga.edu

Fully sequenced

Insert Length: 1800 Std Error: 0.00.

Location/Qualifiers

1..1800

/organism="Magnaporthe grisea"

/mol\_type="mRNA"

/strain="CP987"

/db\_xref="taxon:148305"

/clone="RCW26"

/sex="Mat1-1 hermaphrodite"

/tissue\_type="Mycelium"

/dev\_stage="Day 5 post-inoculation"

/clone\_lib="RCW Lambda Zap Express Library"

/note="Vector: pBluescript excised from Lambda Zap

Express; Site 1: EcoRI; Site 2: XhoI; Day 5

post-inoculation RNAs prepared from Magnaporthe grisea

grown at 23C in the dark with constant gyratory shaking

100 rpm in Vogel's minimal medium containing 0.5% isolated

rice cell walls as the sole carbon source. Library

provided by Sheng-Cheng Wu. Sequences were processed by

one of two methods. Where a full-length alignment to the

M. grisea genome sequence was available, the EST sequence

was trimmed according to the alignment, otherwise sequence

quality was assessed using phredPhrap version 991019 and  
 trimmed according to phd files (0.05) and for vector  
 seqs."

## ORIGIN

Query Match 5.3%; Score 81.4; DB 9; Length 1800;  
 Best Local Similarity 49.7%; Pred. No. 0.0041;  
 Matches 373; Conservative 0; Mismatches 341; Indels 37; Gaps 5;  
 QY 274 GAGACTGCTCACCATCAACGTTGGTGGCGCCGCGGACCAAGGGGGCGCCAAACCTC 333  
 Db 408 GAAGACTGTCTTTTCTCTCGACATCTACGCGCGTCCAAATGCCACTTTCATCTCCAAGTC 467  
 QY 334 CCGTTCATGCTCTGATCTTTGGCGTGGTGTGAGATCGGAGCCCCACCATCTTCCT 393  
 Db 468 CCGTGTGTGATGTTTATCCAGGGCGGGGTCAACAAAATTCAAACTCCAACTACGAC 527  
 QY 394 CCGGCCAGATGTTCAACAGAGTGTCTCATGGCAAGCATCATCCAGTGGCCCTCA 453  
 Db 528 GGCACAGGGCTGGTCAAG-----GCCGGCATTTACGGCTGATGCCATATCTTCA 579  
 QY 454 ACTACGTTGTGCTCTGTTGGGGTCTTGTGATGACATCAAGGCCGAGGCGAGCG 513  
 Db 580 ACTACCGGTCGGGCGCTACGGCTTCTCAC-----CAACGGCAACGAGTGTGAGC 630  
 QY 514 GGAACGCGGCTTGAAGGACCGGTTTGGCATGCGTGGTGGGAGCAACAATTCGCG 573  
 Db 631 CCAACACGGGCTTAAGGACCGGCAAGCGCTCCAGTGGATCAAGAAGAACAATTGCGC 690  
 QY 574 GGTTCGGCGGCGACCCGAGCAAGGTGACATCTTTGGCGAGGGCGGAGCAGTGTCCGTGTT 633  
 Db 691 AGTTTCGGTGGCGACCG-----GACCACGTCGTCCTGATCGCGGTTTCGCGCGGAGG 742  
 QY 634 GTCCACCTCATCTTGAAAGCAGCGGCAACAACAGTCAAGGGCAAGCCGTTGTTCCGCGC 693  
 Db 743 GCGAGCATCTCCCTGCTGATCGGTACGCGCGCAAGGAGGAGGCGCTGTTCCACGG 802  
 QY 694 GGCATCATGCGAGGAGCCATGTCGCGGACCCGCGTGGAGCGGACGACGACGACGAT 753  
 Db 803 CGCAGCGGCGACCGCGGTTGCTGCGACGCGTGTGACCGGTGGAGAGTCCCACTACCA 862  
 QY 754 CTACGACCTCTTTGTCGAGTGTGCTGTCGAGCGCGCAGCGCAAGCTCCGTCGCTT 813  
 Db 863 GTACGACAACTTTGCCATCTTCGAGGCTGTGGGTGCG---CGACTCGTGGCTGCTT 919  
 QY 814 GGCAGTGGCAGCGACACCTTGTGATGCCAACCAACACTTCCTGGGTTCTTGGCGTA 873  
 Db 920 GCGGTCCAAAAACACCCAGCTGCAGCGCGCCCAACAGCGGCTTACCCGGGGC 979  
 QY 874 CTCCTCGTTTGGTGTGCTCCCGGCGCC-----GACGGCAAGACATCACCGATGA 924  
 Db 980 CTCGAGGCGCGCTTGTATCATGTGAACCCCGTGTGACGCGCATGCTCAGCGAGCT 1039  
 QY 925 CATGTACAAGTTGTCGCGCAGCGCAAGTATGCAAGCGTTCCCGTGTATTCATTTGGCGACCA 984  
 Db 1040 GACCTATACCGCTTCGAGGAGGAGGTTCTCTCGCGTTCGCGTCACTTCGCGACGTC 1099  
 QY 985 GAACGACGAGGCAACATCTTTGGTCTTGA 1015  
 Db 1100 AACCAACGACGGCGGCTCTTTGCCCTCGA 1130

## RESULT 15

CG811743/c

LOCUS

DEFINITION

FSAA527TF LargeInsertGenomicLibrary Fusarium virguliforme genomic

clone KMV5F5, genomic survey sequence.

ACCESSION

CG811743

VERSION

CG811743.1

KEYWORDS

GSS.

SOURCE

ORGANISM

Fusarium virguliforme

Fusarium virguliforme

CG811743 781 bp DNA linear GSS 13-NOV-2003  
 FSAA527TF LargeInsertGenomicLibrary Fusarium virguliforme genomic  
 clone KMV5F5, genomic survey sequence.

CG811743.1 GI:38265217

GSS.

SOURCE

Fusarium virguliforme

Fusarium virguliforme

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

## FEATURES

source

Fri Aug 6 10:53:09 2004

REFERENCE 1 (bases 1 to 781)  
AUTHORS Meksem,K., Ishihara,H., Koo,H., Shultz,J., Ali,S., Iqbal,J.,  
Lightfoot,D.A. and Town,C.D.  
TITLE End sequencing of BACs from a fingerprint physical map of the  
causative agent of soybean sudden death syndrome, *Fusarium  
verguliforme*  
JOURNAL Unpublished (2003)  
COMMENT Other GSs: FSAAS27TR  
Contact: Chris Town and K. Meksem  
The Center of Excellence in Soybean Research, Teaching and  
Outreach, Southern Illinois University at Carbondale and Plant  
Genomics, The Institute for Genomic Research  
Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62901-4415,  
USA and 9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 618 453 3103 and 301-838-3523  
Fax: 618 453-7457 and 301-838-0208  
Email: meksem@siu.edu; cdtown@tigr.org (URL:  
http://Fusariumvirguliform.siu.edu)  
Seq primer: TGTAAACGACGGCCAGT  
Class: BAC ends.

FEATURES Location/Qualifiers  
source 1..781  
/organism="Fusarium virguliforme"  
/mol\_type="genomic DNA"  
/cultivar="Monticello"  
/db\_xref="taxon:232082"  
/clone="KMFV5F5"  
/clone\_lib="LargeInsertGenomicLibrary"  
/note="Organ: Hyphae; Vector: pINDIGOBACS; A single spore  
derived culture was used. Hyphae were grown in an  
incubator for four days. Nuclei were isolated and embedded  
in agarose, restriction digested with Hind III. Large size  
DNA fragments were ligated in vector pINDIGOBACS and  
electro-transformed into DH10B cells."

ORIGIN

Query Match 4.9%; Score 75.2; DB 29; Length 781;  
Best Local Similarity 54.7%; Pred. No. 0.0038;  
Matches 198; Conservative 0; Mismatches 153; Indels 11; Gaps 2;  
QY 457 ACCGTTGTCCTCGTGGGGTTCTTGGCTGGTGATGACATCAAGGCCGAGCGCGGGA 516  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
371 ACCGACTCTCATCTTGGGGCTTCATGGCCAGCAGGACATTTGGACGGGGCATCGGCA 312  
QY 517 ACGCCGGCTTGAAGACACAGCGTTTGGGCATGCGAGTGGTGGCAGACACATTGCGGGT 576  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
311 ACCTGGCCTCAAGACACAGCGCATCGCCCTGCCTGATCAAGGAAACATTGCTGCT 252  
QY 577 TCGGGCGGACCGGAGAGGTGACATCTTGGCGGGCGGCGAGCATGTCGTTGTG 636  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
251 TTGGCGGGACCGCTCCCAAGGTACCATCTGGGGGAGAGTGCCGGAGG-----CGG 200  
QY 637 CCACCTCATCTGGAACGAGCGGGACAAACACGTACAAGGGCAAGCCGTTGTCGCGCGGG 696  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
199 CAAAGTTGATACCGAGCCACAGCCTTTGGAGGCAAGACGAGGGTCTCTTCGTGGAAT 140  
QY 697 CATCATGCAAG---GGAGGCATGTGTCGGACCGCGTGGACGGCACGTACGGCAACGAGAT 753  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
139 CATTGCCAGTCTGGCGCTGATGGAACATGACATGAAGAACTTGACCAACCCCTCAGAAGCG 80  
QY 754 CTACGACCTTTGCTCGAGTCTGCTGTGGCAGCGCGGCGGCAAGCTCGCGTGCCT 813  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
79 CTACGATACCATTTCAAGCGGTTGGTGGCAGACAGAGATTAAGTTGCATGCCT 20  
QY 814 GC 815  
Db ||  
19 GC 18